

Obstacles to biodefence

The United States is expected soon to establish its Department of Homeland Security. The precise threats facing the nation are uncertain, but challenges in its preparedness to deal with attacks on health and agriculture are all too obvious.

In late 1999, US, European and Japanese semiconductor and information-technology businesses were severely damaged, losing revenues and profits and raising prices to compensate for a sudden shortfall in the availability of semiconductors and other components. The cause was an earthquake in Taiwan, which not only led to a tragic loss of life, but also disrupted suppliers of 10% of the world's silicon processing capacity and more than half of the world's semiconductor manufacturing, and suppliers of 80% of the motherboards used in computer manufacturing. Miraculously, Taiwan recovered its capacity to export within two weeks. This episode, however, illustrates the exposure of countries to remote sources of disruption, amplifying the obvious point that the ways by which any country could be threatened by terrorists are innumerable.

One year on from the attacks on the World Trade Center and the Pentagon, the United States is on the threshold of establishing a new federal-government Department of Homeland Security (see page 10). As discussed at a meeting organized by the Johns Hopkins Center for Civilian Biodefense Strategies in Baltimore, Maryland, last month, the lack of clarity of possible threats is only one of the problems facing the new department. The meeting's participants, who hold senior positions in the key agencies and the research community, emphasized that the challenge to the department from unpredictability is compounded by a woeful lack of national preparedness for biological attacks, both natural and artificial. Can the world responsibly go on assuming that non-biological terrorist threats are more probable?

The potential dual use — therapeutic and destructive — of developments in biological understanding is all too apparent. It highlights the need to focus all the more on basic biomedical research. Knowledge of the genomes of strains of anthrax and structures of its receptors poses little threat in practice, but understanding the immune system is another story. We are not far away from determining critical nodes in the molecular networks of immunity and its control, and of other key biological circuits. Understanding these will be crucial for overcoming diseases but might greatly simplify the tasks of terrorists. Such networks could be attacked not only by modified pathogens but also by small molecules: biological and chemical weapons should be treated as distinct but parallel paths in the development of defences.

Lack of understanding

The Bush administration has recommended that billions of dollars be distributed across the federal budget for 2003 to combat terrorist threats. Turf wars and battles between politicians over the regional allocation of funds will abound, and champions of various technologies will compete with each other within and beyond the Department of Defense. According to participants at Baltimore — many of whom have considerable experience of working with military staff — one of the greatest problems is a mind-set in defence circles that is steeped in a history of dealing with nuclear threats. There is a persistent lack of appreciation of just how different biological attacks are to traditional threats, compounded by a lack of biological literacy and awareness.

The capacity of countries to respond even to natural infections is weak. In particular, the US Department of Health and Human

Services' ability to acquire vaccines was described by one informed participant as a failure. Moreover, vaccine suppliers come in many nationalities, so here, as in many other aspects of bioterrorist threats, international coordination poses a further set of challenges for which no adequate framework is in place. Again, the leadership in vaccine provision suffers from a lack of scientific understanding. There has not been enough speaking out by learned societies on this issue.

The lack of relevant pharmaceutical capacity could be seen as a sign that the market is incapable of catering for vaccine provision for healthcare, let alone for biodefence. Yet in the longer term, the threat of disruption to business — as illustrated by the Taiwanese earthquake — suggests that there is an underlying market incentive. Could that be turned into investment in a new biodefence industry? One can only be sceptical in the face of the enormous range of possible threats. And the incentives are even weaker in the agricultural biotech industry, where profit margins are smaller than in health, and in which government appears to be even less prepared.

Rapid detection

A lack of preparedness is also mirrored in a lack of sophistication in specifying and maintaining the defensive technology to operational standards, specifically in diagnostics and sensors (see page 8). As described at the meeting, for decades, and especially over the years since the Gulf War ended in 1991, the US government appears to have been deluded in its optimistic faith in the ability of current sensor technologies to help protect the population. Specifications that need to be met are weak, and claims by suppliers are rarely adequately scrutinized.

One priority over the next five years must be to develop a capacity to deliver prompt detection and diagnosis following a chemical or biological attack, along with an ability to mobilize protection and treatment quickly. The latter requires a mix of private and public industrial investment incentives to be developed with high priority, so that pharmaceutical companies would each have two or three related programmes under way, and tens of biotechnology companies would also be developing products. Easily said, but the additional investment required would amount to much more than a billion dollars over several years.

One way of improving scientific and technical awareness of the problems inherent in biological defence, and developing appropriate capacity to respond, is to have key stakeholders collaborate in the development of scenarios of imaginary attacks, however speculative. This process will have the additional benefit of putting under scrutiny the performance of technologies and improving the standards to be applied in manufacture, as well as the ability of leaders to summon up the assistance and information they require in prioritizing their response. These studies would need to include attacks using smallpox or anthrax aerosols, and botulinum toxin in food. Scenarios of this kind have been studied on occasion, and the process can drown in a plethora of competing assumptions and models, so it needs good management. But the lack of national preparedness and high-level appreciation of biological threats suggests that there is more to be done in using such an approach. ■





War of words
Biologist licks his wounds after stem-cell débâcle
p4



Looking up
Namibian telescope sees Africa join the hunt for cosmic rays
p6



World view
Science has its say at Johannesburg conference
p7



Sold short
Original yardstick shows the metre is not long enough
p8

Geneticists lay foundations for human transcriptome database

David Cyranoski, Tokyo

Move over genome, here comes the transcriptome. Last week, 120 researchers from around the world gathered in Tokyo to assemble the core of a transcriptome database, which they hope will one day hold all of the expressed sequences in the human genome.

The database, which should be up and running by December, will be a universal resource for biological research and drug discovery, say the meeting's organizers. "We want to know exactly where the genes are and what they do," says Sumio Sugano, a researcher from the University of Tokyo's Institute of Medical Science.

As the first step in producing proteins, information in genes is transcribed into messenger RNA (mRNA). This process separates the coding sequences of genes from the rest of the genome — often called 'junk' DNA. The transcriptome is the complete set of transcribed mRNA. For years, researchers have studied these transcripts in the form of complementary DNAs (cDNAs), which are made using the mRNA taken from cells as a template. cDNAs represent the mRNA present in the cell, but they are much easier to work with than mRNA itself.



Expressed interest: researchers gathered in Tokyo to assess the integrity of cDNA data.

Now researchers want to incorporate the sequences of all of the human cDNAs into a single database, to be run by the Japan Biological Information Research Center in Tokyo and the DNA Data Bank of Japan (DDBJ) in Mishima. At the Tokyo meeting, researchers analysed cDNA data representing over 20,000 genes — covering more than half of the transcriptome — for inclusion in the database.

Trying to find genes within the human genome sequence often means guessing at



which parts are expressed by looking for certain patterns in the sequence. cDNAs, made from mRNA expressed in cells, offer a more direct route. "This will be a real human-gene catalogue — not predicted from the human genome sequence. These are real transcripts," says meeting organizer Takashi Gojobori, director of the DDBJ in Mishima.

Most of the cDNAs are already publicly

Bush urged to boost funding for physical sciences

Geoff Brumfiel, Washington

The main scientific advisory panel to the White House has joined a chorus of calls for more research funding for the physical sciences in the United States.

In a letter soon to be sent to President George W. Bush, the President's Council of Advisors on Science and Technology (PCAST) will call for a funding increase in fields such as physics and engineering, to match the five-year doubling of biomedical research at the National Institutes of Health, which will be completed next year.

"All evidence points to a need to improve funding levels for physical sciences and certain areas of engineering," says a draft of

the letter, which was discussed by PCAST on 29 August and is expected to be formally agreed and sent within a few days.

Because Bush appointed PCAST only last year, and it comprises scientists and engineers friendly to the administration, lobbyists are optimistic that Bush will heed its advice. "I think this letter is a breath of fresh air," says Michael Lubell, director of public affairs at the American Physical Society in Washington DC, which has advocated an increase for the past few years.

Lubell believes that the proposal will influence Bush and the Office of Management and Budget, which sets the president's annual budget — although he

is not sure if it will influence the president's budget request for 2004, which is being drawn up now for release in February. "The 2004 presidential budget will be extraordinarily tight," he says.

Some scientific societies were less happy that the letter failed to mention other disciplines, such as space science and mathematics, whose funding has languished in recent years. "I'm a little concerned about the way they worded this letter," says Samuel Rankin, director of the American Mathematical Society in Washington DC.

But Lubell says he doesn't think that these fields are being left out. "I think they are implicitly included," he says.

available — but many exist as fragments of the complete cDNA. In addition, the lack of proper categorization, and inconsistencies between the various databases, limits the usefulness of the sequences for research.

"The data will be well-defined and quality controlled through the checks and balances of over a hundred scientists," says Ranajit Chakraborty, director of the Center for Genome Information at the University of Cincinnati Medical Center in Ohio.

To create the data set, the researchers mapped 42,000 cDNAs, collected from six databases around the world, to some 23,000 different regions on the human genome. The overlap of many cDNAs at the same regions will shed light on one of the mysteries of the genome — how so few genes can make the range of proteins that carry out the many functions in human development, and also produce so much variety in people's genetically determined features.

One explanation is that the genes undergo alternative splicing, whereby various mRNAs are produced from the same genomic sequence. By looking at many slightly different cDNAs that cover the same gene regions, researchers say that they will find many examples of these alternate forms of mRNA.

The meeting also offered a large data set, and a platform for debate, concerning non-coding RNA, which does not make protein. Some researchers believe such non-coding RNA has a major role in regulating gene expression, but the idea remains controversial (see *Nature* **418**, 122–124; 2002).

♦ www.jbirc.aist.go.jp/index_E.html

♦ www.ddbj.nig.ac.jp

Name-calling gets stem-cell researcher into hot water

Carina Dennis, Sydney

If you ever find yourself called from the lab bench to testify on a contentious topic, here's a cautionary tale: a prominent stem-cell researcher is this week licking his wounds after being accused of misleading the Australian public over the potential of embryonic stem-cell research.

Alan Trounson, who directs a developmental-biology centre at Monash University in Melbourne, has been widely denounced in the news media for a technical inaccuracy made during briefings to members of parliament. Opponents of the research rounded on the error as evidence that scientists were wilfully misrepresenting their findings.

"As a scientist whose integrity has been put at stake by people with an axe to grind, it has been terrible for me, and the impact on my family has been awful," says a drained Trounson, who was hospitalized last week for a heart condition.

Trounson showed parliamentarians a video of the recovery of a paralysed rat following injection of what he said were human embryonic stem cells. But the cells were embryonic germ cells — strictly speaking, cells from the parts of 5–9-week-old embryos and fetuses from which eggs and sperm later form. Embryonic stem cells are derived from embryos that are just a few days old. The video was taken at the lab of John Gearhart, a biologist at Johns Hopkins University, Baltimore, but the experiment's results have yet to be published.



Tongue-tied: Alan Trounson's attempts to simplify science for parliament backfired in the media.

Although Trounson acknowledges the error, he defends his actions. "Embryonic germ cells had never been explained to these parliamentarians before, so I simplified and just called them embryonic stem cells," he says. "This is not absolutely correct, but they are embryonic and they are stem cells and you can't tell the difference between them."

Several biologists have jumped to Trounson's defence. But he advises others to step gingerly when providing information to politicians. "Be very careful about taking these matters forward unless you get a professional organization to help you, in terms of handling the media and the politics," he says. "You just can't do it yourself."

Australia is currently looking at legislation to regulate human cloning and stem-cell research. Last week, parliament voted to pass a blanket ban on human cloning and to defer the vote on embryonic stem-cell research until later this month.

Dispute over first authorship lands researchers in dock

Alison Abbott, Munich

Rows about the order of authors' names on a research paper can get stormy at times — but they don't often end up in court. That's what has happened, however, in a dispute between two molecular biologists at Germany's University of Göttingen.

The paper in question — an analysis of a tyrosine kinase receptor of relevance to cancer — had been jointly prepared by Marco Ledwon, a former PhD student at the university, and Frauke Alves, the leader of the research team, for submission to the journal *Biological Chemistry*.

But Ledwon, who now works for Lower Saxony's Association of Libraries, took legal action in February, before the paper was submitted, alleging that Alves had substituted her name for his as first author on the final draft without reasonable cause.

The state court of Lower Saxony slapped an immediate injunction on Alves to prevent her submitting the paper for publication



until the court had made a considered ruling.

A few weeks ago, the court ruled in favour of Ledwon. But it did not base its ruling on the relative intellectual contributions of Ledwon and Alves to the paper. Instead it said that the original verbal agreement that Ledwon should be first author had not been disputed in the 14 months during which the paper was being prepared. This understanding constituted an implicit contract, the court said, which Alves had broken.

According to *Laborjournal*, the magazine that first reported the case, Alves contends that Ledwon made insufficient research contribution to warrant first authorship. But Ledwon says he carried out experiments independently and helped to write the paper. The paper has not yet been submitted for publication and the University of Göttingen declined to comment on the case.

Court judgement opens door for study of ancient skeleton

Rex Dalton, San Diego

In a major victory for US archaeologists and anthropologists, a court has ruled that they can go ahead and examine Kennewick Man, a 9,200-year-old skeleton found six years ago in a river-bed in Washington state.

A US district-court judge in Portland, Oregon, ruled on 30 August that the Department of the Interior had been unjustified in its September 2000 decision to hand the skeleton to five Native American tribes, who did not want any research to be done on it.

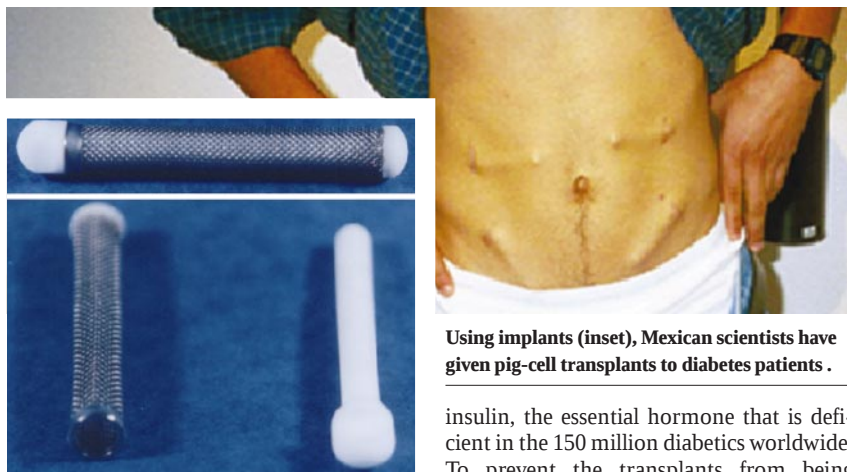
A group of scientists first filed a lawsuit in 1996 to demand the right to study the skeleton, which could provide significant information about early human migration in the Americas. One of them, Robson Bonnichsen, an anthropologist at Texas A&M University in College Station, said: "The decision is terrific, it will benefit everyone."

A spokeswoman for the Department of Justice said that the government wouldn't comment until it had reviewed the decision. But spokesmen for the Native American tribes expressed disappointment, saying that the court had prevented them from exercising their rights under the Native American Graves Protection and Repatriation Act.

The ruling is likely to have implications for the study of early human migration in the United States as it will make it harder for Native Americans to use the act to prevent the study of other specimens. One example is a skull unearthed in 1999 in a bog near Houston, Texas. Radio-carbon tests suggest that it is about 11,000 years old. Such specimens are now more likely to be fully studied, although Native American tribes will still be able to claim the remains and rebury them afterwards.

♦ <http://friendsofpast.org>

Diabetes trial stirs debate on safety of xenotransplants



Using implants (inset), Mexican scientists have given pig-cell transplants to diabetes patients.

Erika Check, Washington

The use of pig cells to treat Mexican patients with diabetes has sparked a fresh round of debate on the ethical and social issues surrounding xenotransplantation.

In the study, Rafael Valdés of the Children's Hospital of Mexico in Mexico City implanted pig cells into 12 adolescents in an effort to cure their diabetes. His approach — which used specialized cells to protect the transplants from the host patients' immune systems — has never been tried before in humans. On 27 August, Valdés told the XIX International Congress of the Transplantation Society in Miami, Florida, that his preliminary results are promising, although other scientists are reserving their judgement.

Valdés' approach has not been properly tested in animals, critics say. They want to know if any benefits from the treatment outweigh the risk of exposing patients — and the rest of the population — to new pathogens that might be carried by the pig cells.

But Valdés says that the Mexican government approved his trial. His critics point out, however, that Mexico has no published guidelines on clinical trials in xenotransplantation.

"I don't think anyone is trying to say to people in other countries, 'We want to tell you what to do'," says David Cooper, an immunologist at Massachusetts General Hospital and past president of the International Xenotransplantation Association. Nonetheless, Cooper is worried by the theoretical potential of xenotransplantation to transfer infection into the human population. He thinks that any proposed approach should at least be proved to work in non-human primates before it is tried out in people.

Valdés gave his trial participants implants of pancreatic cells from pigs, which produce

insulin, the essential hormone that is deficient in the 150 million diabetics worldwide. To prevent the transplants from being rejected, Valdés also implanted cells from the pigs' testicles, known as Sertoli cells. These cells have been shown to protect transplanted tissues from a recipient's immune system in some animal models.

Valdés said in Miami that the implants have allowed one child in the trial to stop taking insulin injections completely, and five others are taking less insulin than before. The result was achieved without using immunosuppressive drugs, which have harsh side effects but are normally given to transplant recipients for the rest of their lives. If confirmed, these results would be unprecedented, researchers say — nobody has ever weaned a diabetic patient off insulin injections without using lifelong immunosuppression.

The results have prompted calls from some researchers for Valdés to study his technique in non-human primates as soon as possible, but Valdés says that he plans to go straight ahead with a trial in 24 more humans.

Critics are challenging Valdés over what steps he is taking to stop his patients contracting infections from pigs. He says he is following US Food and Drug Administration (FDA) guidelines for xenotransplantation trials. Scientists from developing countries often use these guidelines even though the FDA has no authority outside the United States.

But other researchers say that the FDA would not have approved Valdés' trial for several reasons, including the lack of preclinical studies and the fact that his work was in patients too young to give proper informed consent.

"If you have a technique like this that you wouldn't do in Britain or the United States, but it works, where does that leave your ethics?" asks Alan Colman of Singapore-based company ES Cell International. "There is no moral high ground here — but I do think this experiment went beyond the pale."



A plastic cast of Kennewick Man: researchers will soon be able to study the real thing.

Hunt for cosmic rays offers scope for Africa

David Adam, London

As the World Summit on Sustainable Development was set to wrap up in Johannesburg this week, astronomers are heading to southern Africa for the opening of a γ -ray telescope in Namibia. And they hope that the project will serve, in its own small way, as an example of how to build bridges between science in Europe and Africa.

The instrument is the first of an array of four telescopes — called the High Energy Stereoscopic System (HESS) — that will indirectly observe the Galactic sources of cosmic rays. The telescopes will search for faint light called Cerenkov radiation, which is emitted when γ -rays pass through the Earth's atmosphere.

Cosmic rays release large numbers of γ -rays early in their life, so tracking the γ -rays could help researchers locate the source of cosmic rays (see News Feature on page 12). In contrast with the charged cosmic rays, γ -rays are uncharged and always travel through space in a straight line, making them a useful surrogate for the observation of cosmic rays.

But as well as shedding light on a range of extreme cosmic environments, such as



Watching brief: the HESS telescope will study cosmic rays by tracking the γ -rays that they release.

supernovae, the new facility, which is located about 100 km southwest of Windhoek, could boost opportunities for local scientists. The physics department at the University of Namibia in Windhoek is a partner in the 7.6-million euro (US\$7.5-million) project, which is being led by the Max Planck Institute

for Nuclear Physics in Heidelberg, Germany.

Riaan Steenkamp, a cosmic-ray physicist at the University of Namibia, hopes that the arrangement will mean funding for more students in his department. "It's a stroke of luck that they decided to build this in Namibia," he says. "I'm currently the only guy here trained in a relevant field but I hope that within a year or so we can establish a small group."

Student exchange schemes with universities in France and Germany have already been set up, he says, and access to the telescopes could help the university win cash from a new space-science programme in neighbouring South Africa. The telescopes have also sparked national interest — the Namibian Post Office has issued a stamp commemorating its construction, painted by local artist J. J. van Ellinckhuijzen.

In the past, astronomers have been criticized for setting up observatories in the developing world — particularly in Chile — that didn't reach out to local communities, universities or scientists. HESS is being billed as a European–African collaboration by its European sponsors, although the University of Namibia is one of just two partners actually based in Africa, the other being the University of Potchefstroom in South Africa.

"The Europeans were sensitive because of what happened in South America, so they made it a point to involve us," Steenkamp says. Telescope sites in the Southern Hemisphere are popular with astronomers based in the north because they allow them to view different parts of the cosmos, including the centre of our Galaxy.

The first HESS telescope, which was due to be inaugurated on 3 September, should produce its first useful data by the end of the year. The other three telescopes that will make up the array are expected to open by the end of next year.

Norway sinks ocean carbon study

Jim Giles, London

Pressure from environmental groups looks set to scupper an international research team's attempt to test the feasibility of sequestering large volumes of carbon dioxide in the ocean.

The Norwegian government has intervened to block the proposed release of 5 tonnes of liquefied CO₂ off its coast. Three months ago, the researchers' plan to carry out the same experiment on a larger scale near Hawaii was abandoned in the face of environmental objections (see *Nature* 417, 888; 2002).

Now the team — which includes engineers, oceanographers and ecologists from the United States, Norway, Japan and Canada — is running out of options and may have to abandon the idea altogether. That would be a major blow to global efforts to pursue oceanic carbon sequestration as a possible response to global warming — and a big victory for green groups, some of whom regard sequestration as a diversion from the need to cut CO₂ emissions.

The scientists had planned to release the CO₂ at a depth of 800 metres, and monitor its impact on the Norwegian Sea, to investigate whether the ocean could absorb much larger volumes of CO₂ from power plants, for example.

A permit for the experiment was granted by the Norwegian Pollution Control Authority on 5 July. But Børge Brende, the Conservative Party environment minister in Norway's coalition government, decided to review the authority's decision after protests from Greenpeace and the World Wide Fund for Nature. The ministry announced on 22 August that the project would not go ahead.

"We think it should be researched and a decision made on the science," complains one of the researchers, Eric Adams, a hydrodynamicist at the Massachusetts Institute of Technology. "This is politics meddling with science," he says.

The pressure groups argued that the experiment would contravene ocean-pollution treaties. They also object to any large-scale release of CO₂ into the oceans, claiming that it could damage marine ecosystems and would eventually leak back to the atmosphere. Researchers counter that the experiment was designed precisely to investigate whether such fears are justified.

The team is considering what to do next. Adams says they could do the work quite legally in international waters, although most are too deep for this study. The second delay has also complicated matters with the project's funders, as the original time-frame for the experiment has now expired.

Call for cash to end the decay of Berlin's great collections

Quirin Schiermeier, Munich

East Berlin's new government district, with its lavishly renovated ministry buildings, still carries a gaping wound from the Second World War — Humboldt University's crumbling Museum of Natural History. Now, a panel of senior scientists has declared that its 100-million-euro (US\$98-million) renovation is a "national obligation" for Germany.

The wooden windows of the nineteenth-century building — which was bombed by the Allies in 1945 — rattle in the wind, and the plaster on walls and ceilings is falling away, exposing ancient copper water pipes and wiring.

But structural decay is just part of the problem facing what remains one of the world's great natural history museums. So says the report of the panel appointed last year by Humboldt University and the state government of Berlin to advise on how the museum can be revived.

The museum holds over 25 million items, some collected by famous German explorers such as Alexander von Humboldt and the botanist Adelbert von Chamisso. But the collections are threatened by damp, heat and pests, and have not yet been electronically catalogued.

The huge palaeontological, mineralogical and zoological collections are maintained by a staff of 160 and an annual budget of 9 million euros — about a tenth of that of similar museums in London and Paris.

The panel, chaired by Gerhard Neuweiler, a zoologist at the University of Munich and former president of the Wissenschaftsrat, Germany's science council, also suggests a new organization for the museum. Instead of operating its three main departments autonomously, the museum should have a director-general, responsible for integrating its research, exhibits and fund-raising.

But it is unclear where the 100 million euros will come from. The state of Berlin is short of cash and the university's plans to raise funds by selling property have fallen through.

"For the time being, we can hardly expect more than good will," says Neuweiler. At some time in the future he hopes the museum will be saved by money from the federal government.

"The collections belong to our cultural heritage," Neuweiler says. "We owe it to the public and to the international scientific community to conserve them." ■

Gene-bank expansion plan launched at Earth summit

Michael Cherry, Johannesburg

Steps to secure the world's crop diversity and to boost mathematics in Africa are among the initiatives that have emerged from the World Summit on Sustainable Development.

The Johannesburg conference, which ends this week, may have proved difficult for politicians, who were criticized for poor progress on environmental, economic and health issues. But it was useful for the scientists involved in parallel meetings, who discussed food issues and how to build research capacity in developing countries.

One significant move was the launch of an independent Global Conservation Trust. Sanctioned by the UN Food and Agriculture Organization, the trust aims to raise US\$260 million to expand the world's gene banks — collections of seeds and other organic material held to protect crop diversity. The launch followed the release of a report revealing that lack of funds prevents many gene banks from planting samples and harvesting material. "If these collections are allowed to fail, we will lose the valuable crop diversity they contain for ever," says co-author Jeff Waage, an agricultural researcher at Imperial College's Wye campus, UK.

Another initiative, the African Institute for Mathematical Sciences, received a boost when David King, the British government's chief scientific adviser, pledged support. The institute, to be based in the seaside town of Muizenberg near Cape Town, will offer master's degrees in mathematics to students from the whole of Africa. It will be a joint venture by the University of Cambridge, which will supply some of the lecturers, and the South African universities of Cape Town, Stellenbosch and the Western Cape.

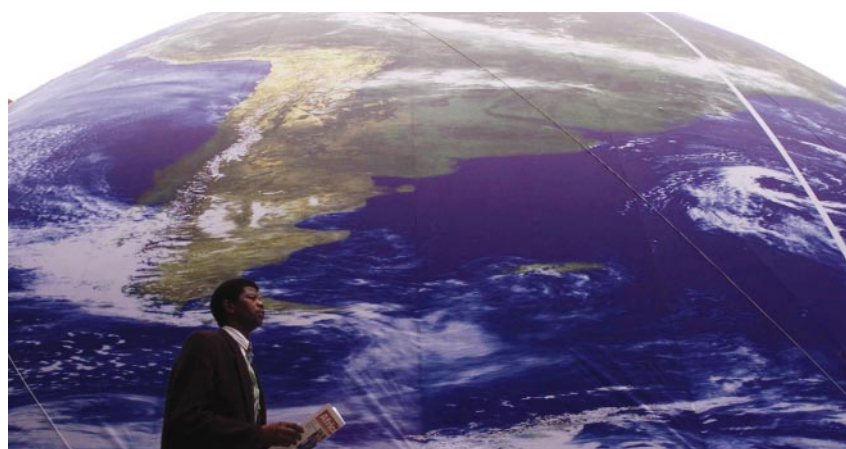
Other debates focused on the potential role of genetically modified (GM) crops in

addressing problems of securing food for the developing world. This included a day-long session organized by AfricaBio, a Johannesburg-based association that promotes biotechnology on the continent.

Supporters of the technology pointed out that GM crops are already used in developing countries. According to Nathalie Moll of the Brussels-based lobby group EuropaBio, 5.5 million farmers grew GM crops last year, more than three-quarters of them small-scale farmers, mostly in China. Others cited health benefits. Mycotoxins, poisonous by-products of the metabolic activity of certain maize fungi, can cause oesophageal cancer, a major killer of black South Africans. These toxins can be reduced by up to 90% in maize engineered to resist an insect called the maize borer, which makes holes in the plant in which fungi grow, botanist Klaus Ammann of the University of Berne in Germany told the meeting.

Many African countries remain opposed to transgenic crops. Zimbabwe and Zambia, for example, recently outlawed their use and have refused to accept GM maize as food aid (see *Nature* **418**, 571–572; 2002). "Biotechnology must not be imposed on Africa, but developed from within it," says Jane Morris of the African Centre for Gene Technologies at the South African Council for Scientific and Industrial Research (CSIR) in Pretoria.

Other delegates pointed out that many countries lack a culture of independent regulatory agencies — a point unintentionally highlighted by Ben Ngubane, South Africa's science minister. He castigated Lungile Shoba, a molecular biologist at the CSIR, for pointing out that the panel set up to discuss transgenic crops contained no opponents of the technology. "Remember I am your minister," he said, wagging his finger at her. ■



One world: and many questions for delegates passing by the giant globe in Johannesburg last week.

MIKE HUTCHINGS/REUTERS

Japan admits to use of biological weapons in Second World War

Tokyo A Japanese court has acknowledged that the country's army used biological weapons against Chinese civilians in the Second World War — but has dismissed compensation claims made by the victims and their families.

The plaintiffs claim that Unit 731 of the Japanese army dropped fleas infected with bubonic plague from aircraft and put cholera bacteria in wells along China's eastern coastline, causing thousands of deaths. They are demanding an apology and ¥10 million (US\$85,000) per person in compensation.

Last week, a Tokyo judge said that medical and military records indicated that Unit 731 had committed the "inhuman acts"; but that the victims were not entitled to compensation under international law. Representatives for the plaintiffs say they will use the admission when they appeal to a higher court.

US steps up biodefence with urban warning system

Washington Two US cities could be equipped with prototype systems for monitoring for the use of biological weapons by 2004, the

US Department of Defense announced on 27 August.

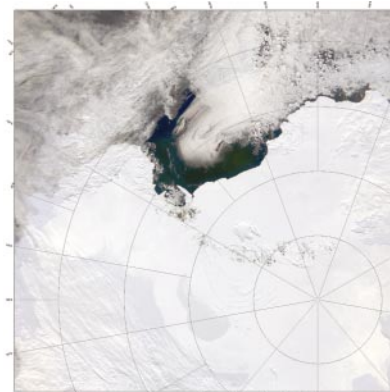
The \$211-million Biological Defense Initiative intends to develop an urban monitoring system that can collect information from numerous sources, such as aerosol monitors, and detect biological agents. The project will be carried out by the Defense Threat Reduction Agency — the section of the Pentagon devoted to countering weapons of mass destruction.

A separate project will examine how military and civilian healthcare data can be used to generate early warning of acts of bioterrorism.

Researchers chill out to tackle dark energy

Washington The poorly understood phenomenon of 'dark energy' is to be studied by a new observatory at the South Pole.

An 8-metre telescope will be used to measure the temperature of the cosmic microwave background (CMB), the sea of photons left over from the early Universe. It will be located at the US National Science Foundation (NSF)'s South Pole Station, a site chosen for its clear, dry skies. Studying the CMB temperature should help scientists to understand how clusters of galaxies form, as gas within the clusters can alter the



Ice work: clear air makes the South Pole the ideal location for studies of dark energy.

temperature of CMB photons. It should also reveal more about dark energy, which appears to be accelerating the expansion of the Universe and is thought to influence cluster formation.

A team of scientists led by the University of Chicago will build the telescope, with funding of \$16.6 million over five years from the NSF.

Cash cuts threaten US plans for bioterror centres

Washington Plans for new centres of excellence for bioterrorism research will have to be scaled down if the budget of the National Institute of Allergy and Infectious

Diseases (NIAID) is cut, the institute's director Anthony Fauci said last week.

The Bush administration had wanted the institute to receive \$3.9 billion for the fiscal year 2003, more than \$1 billion of which was to go towards the construction of new biosafety labs, basic bioterrorism research, development of therapies and vaccines, and clinical research. But in July the Senate committee that sets budgets for the National Institutes of Health took \$263 million out of the NIAID's budget for that year and redistributed it to other institutes, including the National Institute for Bioimaging and Bioengineering.

The NIAID had planned to fund up to five regional centres of excellence for bioterrorism and emerging infections, but would probably have to drop to two or three centres if the cuts are implemented, said Fauci. The House and Senate must agree on final spending before the bill becomes law.

Britain told to track transgenic trade

London The international movement of cloned and genetically modified animals should be monitored to prevent smuggling, maintain consumer choice and protect the environment, the British government has been told.

The Agriculture and Environment Biotechnology Commission (AEBC), a UK government advisory body, says that the arrangements must be in place before the production and trade of transgenic animals becomes widespread. Current movement regulations do not require exporters to state if animals or reproductive material have been modified or cloned.

Animals and Biotechnology, published on 3 September, says the Cartagena Protocol on Biosafety could provide a suitable starting point. The protocol requires countries exporting transgenic animals or plants to obtain the informed consent of importing countries, but it has not yet been ratified by enough countries to enter into force and does not cover cloned animals or embryos.

Book reveals short measure but still the metre rules

London A new book will bring satisfaction to the many Britons and Americans who are reluctant to use the metric system — the metre is apparently shorter than it should be.

The metre was first defined as one ten-millionth of the distance from the Pole to the Equator. But in *The Measure of All Things* (Free Press/Little, Brown), science historian Ken Alder claims that the eighteenth-century French astronomers given the task of



Pierre Méchain:
metre made.

working out the measurement bungled their calculations. As a result, the metre is 0.2 mm short of its original definition.

According to Alder, expedition log books show that one of the astronomers, Pierre Françoise André

Méchain, discovered that

two of his measurements made from the same place did not agree. Letters show he was driven to the brink of madness by his decision to conceal the error. He eventually died of malaria on a trip to correct the measurement.

The second astronomer, Jean Baptiste Joseph Delambre, was aware of the cover-up and admitted in the logs that: "I have not told the public what it does not need to know."

Correction The recent News Feature on immune tolerance (*Nature* **418**, 364–366; 2002) stated that researchers at Dartmouth Medical School had halted a clinical trial of a monoclonal antibody to treat multiple sclerosis when one patient developed blood clots. In fact, the patient who developed clots was enrolled in another trial, which was investigating whether the same antibody could ease symptoms of Crohn's disease.

Mission impossible?

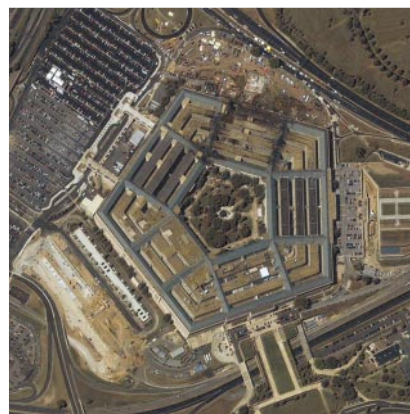
A new Department of Homeland Security is to be given the task of defending the United States against further terrorist attacks. Geoff Brumfiel outlines the challenges facing its research wing.

One year ago, on a crisp and gloriously sunny morning, the United States' sense of domestic invulnerability was shattered. Grief, shock and anger over the attacks on the World Trade Center and the Pentagon have since been joined by anxiety about the nation's ability to prevent future outrages. This has been fuelled by media criticism of intelligence agencies' failure to warn of al-Qaeda's assault, and of the government's handling of the subsequent anthrax mailings — for which a perpetrator has yet to be identified.

In an attempt to reassure a troubled public, President George W. Bush unveiled plans on 6 June to upgrade the existing Office of Homeland Security, created in the aftermath of the 11 September attacks, to a full department of the federal government. Research is an integral part of the plan. "Our scientific community is serving on the front lines of this war," Bush told researchers in a speech at the Argonne National Laboratory in Illinois in July. "You all know better than anybody, when we research and we set priorities, this great nation can achieve any objective."

With bills to establish the Department of Homeland Security still being considered by the Congress, the details remain unclear. But the department is likely to have a budget of \$32 billion a year, including several hundred million for research. It will also help to manage the much larger sum requested for biodefence research under the National Institutes of Health (NIH).

Counter-terrorism experts agree that science will be key to addressing the threat. But they warn that the department's research



Devastating: satellite pictures of Manhattan and the Pentagon the day after the 11 September attacks.

wing faces some major challenges. First, it must interact with the department's operational divisions, many of which will be plucked from other agencies with no strong scientific culture. Second, it must coordinate its own activities with other agencies and departments whose research efforts feed into the homeland-security agenda. Above all, it must respond to a nebulous threat. "Everybody moving into this department will require exceptional talents," observes Parney Albright, assistant director for homeland and national security at the White House Office of Science and Technology Policy.

A tricky target

Scientific research has long been central to US national security policy. Throughout the cold war, US researchers strove to develop better weapons and intelligence-gathering technologies than their Soviet counterparts. But when confronting terrorism, where the threats are diverse and hard to assess, it is not so easy to set research priorities. "It's going to be more complex than building a rocket or a nuclear weapon," says Page Stoutland, deputy division leader for counter-terrorism at the Lawrence Livermore National Laboratory in California.

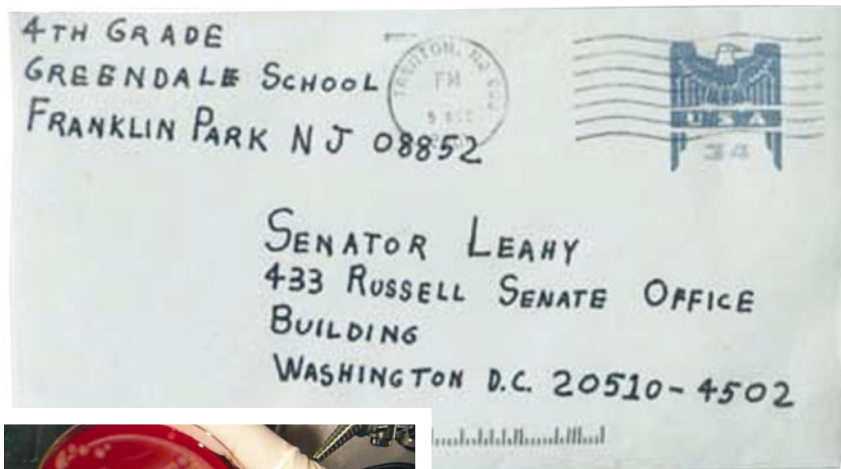
The main concern is the possibility of terrorists gaining weapons of mass destruction. These include a long list of potential chemical and biological weapons — and, even if nuclear weapons are beyond their grasp, terrorists might be able to put radioactive material into a 'dirty' conventional bomb. One key challenge for the department's research wing will be to develop technologies to detect

any attack rapidly, determine its source and respond quickly to mitigate its effects.

According to counter-terrorism experts, the threat from biological weapons is a top priority for research. There are clear vulnerabilities to be addressed, and the promise of developing the means to mitigate attacks. Last year's anthrax mailings, for instance, might have killed many more than their five victims, had the bacterium used been resistant to antibiotics. Although anthrax cannot spread from person to person, other potential bioweapons — the smallpox virus, for example — are communicable. In these cases, the ability to identify rapidly when an attack has occurred, to deploy vaccines, and to isolate and treat infected people will be crucial.

But an enormous amount of work remains to be done. "The development of new vaccines, antiviral drugs and antitoxins is in a pretty sorry state," says Steven Block, a biophysicist at Stanford University in California who is a member of the JASONs, a group of scientists that advises the US government on issues such as bioweapons. Block believes that basic research into disease pathology and immune responses will help to counter a broad range of threats. It may be possible, for example, to develop generic vaccine technologies that could quickly be applied against any agent — even one genetically modified to increase its potency.

"This research's utility goes beyond the next bioweapon attack," agrees Claire Fraser, who heads The Institute for Genomic Research in Rockville, Maryland, which has sequenced the anthrax bacterium's genome and helped to analyse the strain used in the



Breached defences: last year's anthrax attacks revealed clear vulnerabilities needing research.

postal attacks. "Understanding the biology of infective agents, improving diagnostics and determining what makes a great vaccine will have a big impact on treating more run-of-the-mill diseases," she says.

But the need to set priorities in biodefence research illustrates the difficulties facing the homeland-security department. Initially, Bush proposed transferring the \$1.75 billion requested for the NIH for biodefence research in 2003 to the new department (see *Nature* **417**, 675; 2002). But the money is now expected to remain within the NIH, with the homeland-security department having joint management responsibility. The NIH will award grants and the new department will help to set priorities. But how this will work has yet to be decided, as have the department's links with other bodies working in the field, such as the Defense Advanced Research Projects Agency (DARPA).

Divided by defence

Tony Fauci, director of the National Institute for Allergy and Infectious Diseases in Bethesda, Maryland, which will run the NIH's biodefence programme, is optimistic about relations with the new department. "We feel that we are well-positioned at the NIH to do the kinds of research that would provide the Department of Homeland Security with what it needs," he says. But Washington insiders and senior scientists warn that joint-management arrangements are fraught with difficulty. "At this point, I'm sceptical whether it's going to put us in a position where things are handled more efficiently than they are now," says Fraser.

The fact that biodefence research will yield results applicable to infectious diseases in general also raises thorny questions about the dissemination of results of interest both to researchers and terrorists. Against this background, bodies that represent biologists

remain suspicious of the new department's involvement. "We really think the research is dual-purpose and that the NIH would be in the best position to set priorities," says Janet Shoemaker, the American Society for Microbiology's director of public affairs.

Another major task for the department's researchers will be developing techniques to identify terrorists before they commit atrocities. "Once a terrorist moves towards his target, half the battle is already lost," says Magnus Ranstorp of the Centre for the Study of Terrorism and Political Violence at the University of St Andrews, UK. He stresses the importance of developing analytical tools to glean information about terrorist activities from the Internet and other communications networks, and from databases such as those held by financial institutions.

Won Kim, president of Cyber Database Solutions in Austin, Texas, says that this will involve research to modify the 'data mining' programs used by banks and other companies to probe their records. "From databases

of, say, bank transactions and passport controls, this technology can discover unusual patterns that could lead to terrorists," he says.

Refining these tools to comb the vast reaches of cyberspace will be no small task, says David Farber, a computer and information scientist at the University of Pennsylvania in Philadelphia. The challenge, he says, is to develop software that can exceed the capabilities of human analysts in spotting suspicious patterns among reams of data. "It's probably one of the hardest problems in computer science," says Farber.

Here again, the new department's relations with other agencies may present difficulties. Ideally, researchers developing data-mining tools would have access to raw data gathered by the CIA and the FBI. But the proposals currently before Congress do not call for these data to be made available. "There are great sensitivities among existing intelligence agencies," observes William Happer, a physicist at Princeton University in New Jersey who has served on numerous panels advising on civilian and military research.

Lateral thinking

Relations between research and operational divisions will also require attention. The department is to incorporate a diverse range of existing agencies — including the Coast Guard, the Customs Service and the Immigration and Naturalization Service — many of which have no strong tradition of interacting with in-house researchers to assess their technological needs.

Even the structure of the department's research arm remains hazy, for now. Various strategies have been discussed, including launching new research centres at universities and establishing homeland-security divisions within existing national laboratories. Senator Joseph Lieberman (Democrat, Connecticut), whose staff were working on proposals to establish a homeland-security department before Bush adopted the idea, also favours launching a counterpart to DARPA to explore long-term, innovative projects.

But all of the proposals include the appointment of an undersecretary to run the new department's own research programme and to coordinate with other agencies to ensure the relevance of research across the federal government to the security agenda. The person chosen will require formidable talents. "Whoever runs this needs to be a pretty skilled Washington in-fighter," says Happer.

He or she will also need to move quickly. The legislation establishing the department could be finalized as early as this month, and an immense weight of public expectation will be brought to bear from day one. Says Albright: "As soon as that department unlocks its front doors, the American public is going to expect that we are prepared."

Geoff Brumfiel is *Nature's* Washington physical sciences correspondent.



High alert: public expectations for government action are high following last year's attacks.

Let's catch some rays

Particles with hundreds of millions times more energy than those in physicists' accelerators regularly strike the Earth, but no one is sure where they come from. Philip Ball reports on attempts to solve the mystery.

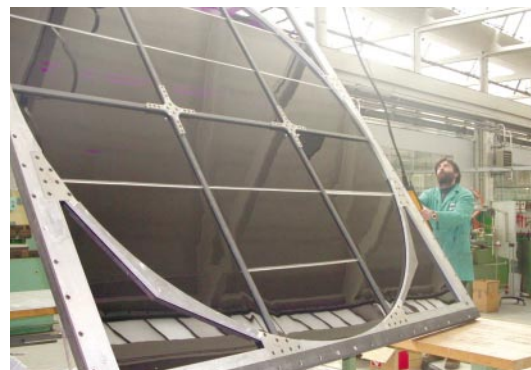
Argentina's Pampa Amarilla desert is filling up with water. Across thousands of square kilometres of the desert's flat plains, engineers are busy building water tanks. By 2005, 1,600 of the 11-cubic-metre tanks will be in place. But the semi-arid grasslands will not reap the benefit — the tanks are there to detect the high-energy subatomic particles that stream unseen into Earth's atmosphere.

The most energetic of these particles — called ultra-high-energy (UHE) cosmic rays — hit the Earth with about 10^{20} electron volts, or 16 joules, of energy. That's about the same as a fast-moving baseball, but carried by just a single subatomic particle. No one knows how the particles gain their energy, or why they do not lose it before arriving at Earth. Some physicists speculate that they are accelerated by shock waves from supernovae, or the sling-shot action of supermassive black holes. Others invoke ideas that lie beyond the current framework of theoretical physics.

The tanks in the desert are part of the Pierre Auger Observatory (PAO), named after the French physicist who first detected cosmic rays from the ground. Most cosmic



Desert manoeuvres: using water tanks (above) and fluorescence detectors (right), researchers hope to detect ultra-high-energy cosmic rays at the Pierre Auger Observatory, currently being built in Argentina (below, left).



rays are protons, although a few are nuclei from elements such as helium and carbon, and some are lone electrons. In 1938, Auger showed that Geiger counters can be used to detect the shower of particles — mainly high-energy photons, electrons and muons, which are heavier cousins of the electron — created when cosmic rays collide with molecules in Earth's atmosphere.

Power shower

In the early 1960s, an array of detectors at Volcano Ranch near Albuquerque, New Mexico, first spotted UHE cosmic rays. They recorded a particle shower that seemed to originate from a cosmic ray with an energy of 10^{20} eV — much more energetic than any previously detected¹. At first, the event was not greeted with great surprise. The flux of cosmic rays was known to decrease more or less smoothly as their energy increases: 10^{12} -eV rays fall over every square metre of Earth's surface every second, and a 10^{19} -eV particle would be expected to strike a square kilometre about once a year. Cosmic rays with energies of 10^{20} eV were not unexpected, but were thought to be very rare. "It was assumed the spectrum would just go on," says John Matthews, who studies cosmic rays at the University of Utah in Salt Lake City.

But a few years later, UHE cosmic rays suddenly became profoundly puzzling. In 1965, physicists discovered the cosmic microwave background (CMB) — a sea of photons left over from the Big Bang. Space is filled with these photons, and they interact with UHE

cosmic rays. Calculations by physicists Kenneth Greisen of Cornell University in Ithaca, New York, and Georgi Zatsepin and Vadim Kuz'min of the USSR Academy of Sciences' Lebedev Institute of Physics in Moscow showed that successive collisions with CMB photons lowers the energy of a UHE cosmic ray until it drops below 6×10^{19} eV, at which point the nature of the collisions changes and no energy is lost^{2,3}. It only takes a few collisions to reach this 'GZK cut-off', so how could the Volcano Ranch detectors have seen a cosmic ray above this threshold?

Experimental error has been ruled out. The Haverah Park observatory near Leeds in northern England detected UHE cosmic rays





Since Pierre Auger worked out how to catch cosmic rays, many detectors, such as the Fly's Eye (right), have spotted high-energy particles.

during the 1970s, and the Akeno Giant Air Shower Array (AGASA) near Tokyo has logged 10 cosmic rays with energies above 10^{20} eV since 1993. In 1991, the Fly's Eye detector at the US Army's Dugway Proving Ground in Utah observed a massive shower of 200 billion secondary particles, apparently produced by a cosmic ray with an energy of 3×10^{20} eV (ref. 4). At five times greater than the GZK cut-off, this is the highest-energy particle ever recorded. There seems no question that these energetic particles exist, but how do they manage to reach Earth?

Violent neighbourhood

The truth must lie in the fact that the GZK cut-off is not absolute. UHE cosmic rays that are formed close enough to Earth would experience fewer collisions before reaching us, so their energies on arrival could still be above the cut-off. Physicists estimate that the source of the UHE cosmic rays must be within 200 million light years of Earth — which means that they come from nearby galaxies.

The problem is that astronomers are not sure which events are sufficiently violent and close enough to create UHE cosmic rays. "The number of sources that are close enough and are capable of producing particles of 10^{20} eV are very few," says Alan Watson, a physicist at the University of Leeds who works on the PAO team. Matthews agrees: "When we look back in the direction that the cosmic ray with 3×10^{20} eV came from, there is nothing there that looks like a potential source."

So what kind of events could generate such energetic particles? The most popular explanations are 'bottom-up' processes, in which lower-energy particles are accelerated to become UHE cosmic rays. Shock waves, perhaps generated when the expanding shell of matter from a supernova collides with the gas and dust between stars, are one possible cause, and evidence for such a mechanism has emerged in the past decade.

During the mid-1990s, the Japanese Advanced Satellite for Cosmology and Astrophysics detected X-rays from a 10,000-year-old supernova. Analysis showed that these X-rays were emitted by high-energy elec-



trons⁵. This April, researchers working on the Japanese–Australian CANGAROO project — the Collaboration between Australia and Nippon for a Gamma Ray Observatory in the Outback — added another link between supernovae and cosmic rays. They reported that high-energy γ -rays, apparently generated by high-energy protons, seem to come from a supernova remnant in our Galaxy⁶.

Both results suggest that high-energy particles of the type seen in cosmic rays are produced by supernovae. But these shock waves can only generate cosmic rays with energies up to about 10^{15} eV. Collisions between galaxies, or the jets issuing from massive black holes that are devouring stars and gas, could create bigger waves. But because there is not yet any direct evidence for either of these mechanisms, other bottom-up explanations are also being considered.

One such alternative involves supermassive black holes. If the black holes are spinning, particles drawn towards them could be pumped back out with very high energies in a kind of gravitational slingshot effect. At April's meeting of the American Physical Society in Albuquerque, New Mexico, Elihu Boldt of NASA's Goddard Space Flight Center in Greenbelt, Maryland, and Diego Torres of Princeton University in New Jersey reported that some of the UHE cosmic rays seen by AGASA have trajectories that line up with four galaxies containing supermassive black holes, all of them within 100 million light years of Earth. Crucially, these black holes are dormant — they are no longer devouring interstellar gas and dust. If they were doing so, they would generate an abundance of photons that would collide with the UHE cosmic rays and drain their energy.

But some believe the apparent alignment could just be a coincidence. "It sounds interesting, but as yet it isn't convincing to me," says Ryoji Enomoto of the University of

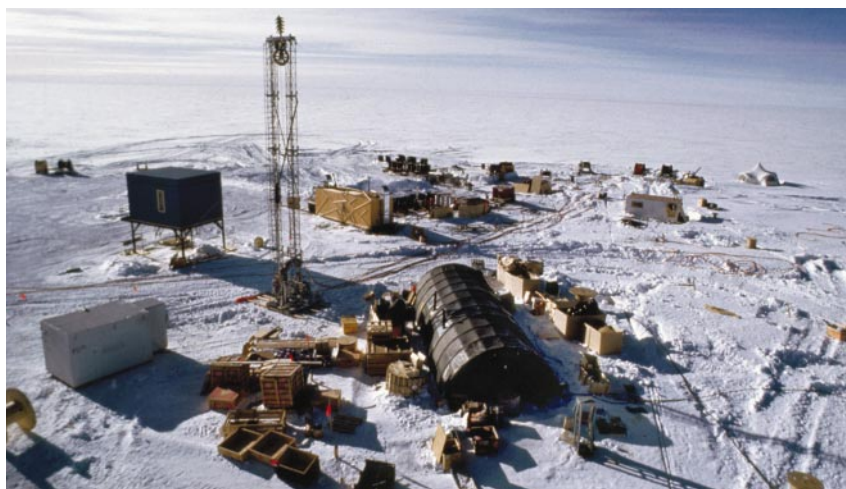
Tokyo, a member of the CANGAROO team. Others dispute that the alignment even exists. "We see nothing like this," says Matthews, who is involved with the successor to Fly's Eye, the High Resolution Fly's Eye. His analysis of the AGASA data indicates that the UHE cosmic rays are uniformly distributed across the sky.

Breaking the rules

Data from the PAO and other observatories should help to settle the issue. In the meantime, more speculative theories are prospering. 'Top-down' explanations propose that UHE cosmic rays are created when even more energetic particles are slowed down. One suggestion is that the particles are produced by the decay of 'topological defects' — wrinkles in space-time left over from the Big Bang. Such defects would create an immense concentration of energy, and their decay would generate new kinds of super-heavy fundamental particles, which could decay into UHE cosmic rays.

This theory lies outside of the realm of established physics, and it is not the only idea associated with UHE cosmic rays to do so. Because the rays have hundreds of millions of times more energy than particles in high-energy physicists' accelerators, their interactions could be used to study areas of physics beyond the limits of our accelerator technology. Jonathan Feng of the University of California, Irvine, and Alfred Shapere of the University of Kentucky at Lexington have proposed that cosmic rays with energies above 10^{16} eV would generate such a high concentration of energy when they collide with atoms in the atmosphere that they could create mini-black holes⁷.

The standard model of particle physics says that UHE cosmic rays should not be energetic enough to do this. But this prediction would change if, as some physicists suspect,



John Matthews says more data are needed to pinpoint the source of cosmic rays. By tracking neutrinos, AMANDA (left) may help this effort.

our Universe has more dimensions than the three of space and one of time. "Gravity is a relatively weak force in our four-dimensional world," says Feng. "But this could be because it gets diluted by spreading out in extra dimensions, unlike the other forces. If that's the case, less energetic particles can create black holes as long as they pass close enough to each other." Such black holes would decay almost instantaneously, but the shower of particles caused by this decay could be detected on the ground. If Feng and Shapere are correct, the PAO will detect around 30 showers a year from mini-black holes.

Space-time wrinkles and other extreme 'top-down' ideas elicit sceptical responses from some researchers. "I tend not to believe that the answer will be new physics," says Watson. But to test these ideas, and to compare different theories of the origin of UHE cosmic rays, researchers badly need more data on the rays that reach the Earth.

Double exposure

The PAO uses two methods to spot UHE cosmic rays, both based on detecting particle showers. Particles moving faster than the speed of light in water will emit a dim blue light, called Cerenkov radiation, as they pass through the water tanks. The particles also produce faint bursts of light when they excite nitrogen molecules in the atmosphere. This 'air fluorescence' is very dim — equivalent to a 20-watt bulb viewed from several miles away — but can be picked up on clear moonless nights by the PAO's second set of detectors: an array of sensitive photomultipliers.

By making dual detections, PAO researchers will have two independent measures of the cosmic rays' energies, allowing better calibration. "An essential feature of the PAO is the hybrid nature of the instrumentation," says Watson. "We can get signals in both detector systems 10% of the time." The data can be used to infer the trajectory of particles in a shower, and hence the path of the cosmic ray. Astronomers can then track the ray's path back and look for potential sources.

About one fortieth of the PAO's array — 40 tanks and 12 fluorescence detectors — is currently up and running. Watson says another 100 tanks and 12 fluorescence detectors should be operational in about 10 months time, and the observatory should be complete by 2005. A northern hemisphere counterpart is also planned, although no site has yet been chosen.

The search for sources of UHE cosmic rays will also soon be bolstered by data from detectors designed to spot high-energy neutrinos. The connection between neutrinos and cosmic rays is still unconfirmed, but some physicists suspect that certain astrophysical processes could produce high-energy versions of both types of particles — so tracking the neutrinos could help to pinpoint the source of UHE cosmic rays.

Neutrinos hardly interact with matter at all, but when they do, they typically produce muons that generate Cerenkov radiation. So water tanks can be used to detect neutrinos, although they have to be buried underground to shield them from the particle showers produced by cosmic rays. Several detector projects have successfully studied low-energy neutrinos over the past two decades, but building a device to snare high-energy versions is more challenging.

The neutrino trail

High-energy neutrinos are far less common than their low-energy counterparts. The two types can be distinguished by the Cerenkov trails left by the muons they create: the higher the neutrino's energy, the farther the muon travels. But high-energy neutrinos can create muon trails that are hundreds of metres long, and once the trail leaves the water tank there is no way of judging where it ends. So for researchers to be confident that they have seen a high-energy neutrino, they need to use very big volumes of water.

The solution is to abandon tanks of water and instead use the oceans or ice-caps, filtering out cosmic rays by burying the detectors deep beneath the surface. This October,

European researchers in the ANTARES project will begin deploying a network of photo-detectors at the bottom of the Mediterranean Sea. Around the same time, another international group will start shipping equipment to the South Pole so that the Antarctic Muon and Neutrino Detector Array (AMANDA), a high-energy neutrino detector that monitors Cerenkov trails in ice, can be upgraded. When completed in 2005, the ANTARES detector will monitor tens of millions of square metres of water. IceCube, the upgraded version of AMANDA, will cover a cubic kilometre of ice by the end of the decade.

Meanwhile, NASA has plans to send the search for UHE cosmic rays into orbit. Its orbiting wide-angle light collectors project, known as OWL, would monitor air fluorescence events from two orbiting detectors, providing a much wider search area than any ground-based detector. It will be on the lookout for cosmic rays with energies in excess of 10^{19} eV, and mission scientists hope to see several hundred events per year with energies of more than 2×10^{20} eV. OWL is still in the planning phase, but researchers are aiming for a launch in 2015.

With new detector projects in the pipeline, researchers studying high-energy neutrinos and cosmic rays face a busy few years. And with such rich theories competing to explain the origin and effect of these particles, it is an exciting time. As the University of Chicago astrophysicist David Schramm, who died in a plane crash in 1997, put it: "It's an interesting field indeed where the most conservative explanations involve super-massive spinning black holes."

Philip Ball is a consultant editor of Nature.

1. Linsley, J. *Phys. Rev. Lett.* **10**, 146–148 (1963).
2. Zatsepin, G. T. & Kuzmin, V. A. *Sov. Phys. JETP Lett.* **4**, 78–80 (1966).
3. Greisen, K. *Phys. Rev. Lett.* **16**, 748–751 (1966).
4. Bird, D. J. *et al. Astrophys. J.* **441**, 144–151 (1995).
5. Koyama, K. *et al. Nature* **378**, 255–258 (1995).
6. Enomoto, R. *et al. Nature* **416**, 823–826 (2002).
7. Feng, J. L. & Shapere, A. *Phys. Rev. Lett.* **88**, 021303 (2002).

Pierre Auger Observatory

♦ www.auger.org

French users need European neutrons

Big projects may not be popular at present, but policy-makers must think of the future.

Sir — As pointed out in your Opinion article “Anyone for neutrons?” (*Nature* **417**, 883; 2002), the European Spallation Source (ESS) user meeting in May was an impressive demonstration of the lively activity of European neutron users. Five propositions for hosting the site of the ESS were presented, with backing from regional and/or national authorities.

It was surprising and regrettable to note the absence of a proposal for a French site. France has a long and deep tradition in the field of scientific use of neutron beams. It hosts the most powerful neutron source in the world, the Institut Laue Langevin (Grenoble), and a modern and well-equipped national source, the Laboratoire Léon Brillouin (Saclay). The French user community has declared its scientific interest in the ESS project on several occasions, and the French Neutron Users Society (www.sfn.asso.fr) has fully

approved the goal of a new-generation neutron source in the next decade. A French team is heavily involved in the design and evaluation of the ESS.

Despite this commitment, no French site was proposed at Bonn, reflecting the obstacles raised by the diverging approaches of its national bodies (the Ministry of Research and Technology, and the two major funding agencies, the CEA and the CNRS). Moreover, the absence of high-level decision-makers representing French authorities at Bonn appeared as a sign of indifference.

It may be difficult in France at the moment to support large research infrastructures, and the financial climate may be unfavourable. But this does not excuse sacrificing the future. One cannot make progress by sticking to fixed schemes of funding and maintaining a conservative outlook on the merits of neutron science.

The Bonn meeting showed the multi-disciplinary and scientific potential of the ESS project. The present sources are needed to assure continuity, and to maintain and complete the training of the scientists who will run and use the source when it is completed.

As is usual in the case of any large-facility project, the construction phase of the ESS would require funding beyond the level of current functioning. But, considering the rapid expansion of the areas of science that will benefit from the new facility, the running costs will surely be reasonable.

F. Leclercq (president)*, H. Mutka (vice-president)†, Société Française de la Neutronique

*LASIR UMR 8516 du CNRS, 13 rue de Toul, F-59046, Lille Cedex, France

†Institut Laue Langevin, BP 156, F-38042 Grenoble Cedex 9, France

Shared knowledge can combat malaria

Sir — Andrew Clark in his excellent News and Views article “Malaria variorum” (*Nature* **418**, 283–285; 2002) discusses the extent of genetic variability exhibited in populations of *Plasmodium falciparum*, the causative agent of the most lethal form of malaria, and the implications for vaccine development and for the emergence of drug resistance. He concludes that a shared resource of *P. falciparum* isolates is needed to address the problem and to advance our understanding of the evolution of this species. We agree.

The Malaria Research and Reference Reagent Resource Center (MR4) was established in 1998, partly to provide the resources available to malaria researchers worldwide that Clark discusses, and is managed by the American Type Culture Collection (ATCC). Under contract to the National Institute of Allergy and Infectious Diseases (NIAID), MR4 makes reagents available free of charge to registered individuals worldwide (see www.malaria.mr4.org). Further information can be found at the NIAID website (www.niaid.nih.gov/dmld/malaria/malrep/default.htm). MR4 currently has more than 450 reagents, including numerous isolates, strains and species of *Plasmodium* and *Anopheles* mosquitoes as well as antibodies, antigens, DNA libraries and primers and microarray chips.

Training programmes, workshops and technology transfer to malaria-endemic regions are also important components of MR4. We encourage researchers to collaborate with MR4 to develop and support the needs of the community, as well as to contribute resources that can thereby be made more widely available.

Yimin Wu*, M. John Rogers†

*MR4/Protistology, ATCC, 10801 University Boulevard, Manassas, Virginia 20110, USA

†National Institute of Allergy and Infectious Diseases, 6610 Rockledge Drive, Bethesda, Maryland 20892, USA

Prokaryote taxonomy online: challenges ahead

Sir — Following calls for online taxonomic databases encompassing all organisms living on Earth^{1–4}, we wish to point out that extensive taxonomic information on the prokaryotes is already available online. The Approved Lists of Bacterial Names, including names and their nomenclatural history, are on a website that is updated bimonthly (www.bacterio.cict.fr). On 9 July 2002 the lists named 5,866 species, classified in 1,104 genera, Archaea and Bacteria combined. Comprehensive information on prokaryote taxonomy is in two handbooks, *The Prokaryotes* and *Bergey's Manual of Systematic Bacteriology*, both published by Springer-Verlag. The former is published online (www.link.springer.de/link/service/books/

10125), and the database of the second edition of the latter (www.cme.msu.edu/bergeys) exists in electronic form, so could easily be made available on the web if the publisher gives permission.

In the case of the prokaryotes, obstacles virtually unknown to the botanist and the zoologist must be overcome to achieve the goals mentioned by Gewin and others. First, the species concept for bacteria is still not well-defined^{5,6}. Second, sequencing of 16S ribosomal RNA genes collected from terrestrial and aquatic environments shows that the number of known prokaryotic species probably forms only a fraction of 1% of the number in nature. Morphological diversity is small, and therefore digital images of type specimens are of little help in online taxonomic databases. The lack of innovative culturing techniques and of well-trained taxonomists to study the isolates are of great concern. Some of the most abundant prokaryotes on Earth defy microbiologists' attempts to isolate them. Here lies one of the major challenges to future taxonomists.

Aharon Oren*, Erko Stackebrandt†

*Institute of Life Sciences and Moshe Shilo Minerva Center for Marine Biogeochemistry, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

†Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroderweg 1b, D-38124 Braunschweig, Germany

1. Godfray, H. C. J. *Nature* **417**, 17–19 (2002).

2. Lee, M. S. Y. *Nature* **417**, 787–788 (2002).

3. Gewin, V. *Nature* **418**, 362–363 (2002).

4. Bisby, F. A. et al. *Nature* **418**, 367 (2002).

5. Cohan, F. M. *Syst. Biol.* **50**, 513–524 (2001).

6. Stackebrandt, E. et al. *Int. J. Syst. Evol. Microbiol.* **52**, 1043–1047 (2002).

A feast for the mind

An introduction to the concept of mind provides plenty of food for thought.

Exploring Consciousness/ Consciousness

by Rita Carter

University of California Press/Weidenfeld & Nicolson: 2002. 319 pp. \$34.95, £20

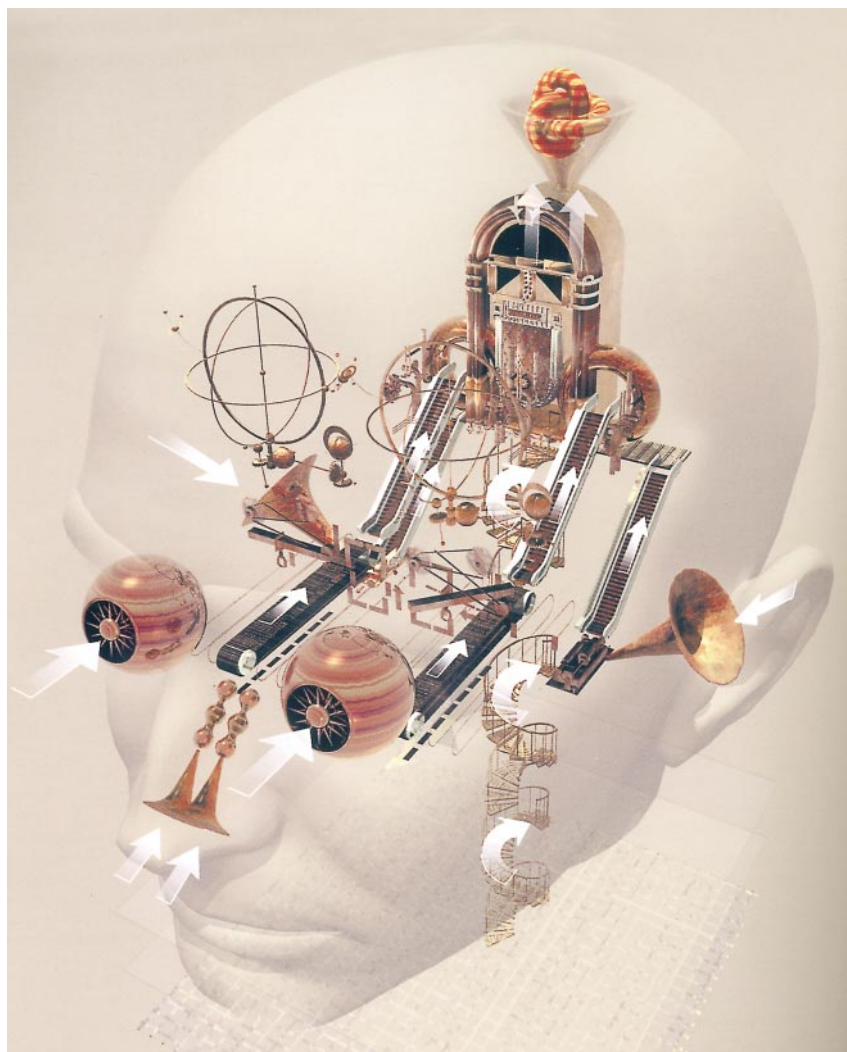
Rosaleen McCarthy

What makes a successful dinner party? The guests? The food? The conversation? I am sure that Rita Carter would host a wonderful party — at least judging by her books. She is an award-winning scientific journalist and author who conveys a real sense of excitement and fascination with her subject matter. She would certainly know what to put on the menu, who should be on the guest list, and how to guide the conversation into interesting areas. However, few would confuse a dinner-party conversation with an academic seminar, and *Exploring Consciousness* should not be confused with an academic text. It is instead an entertaining introduction to a complex subject, written in a lively and accessible manner, and is likely to capture the imagination of a non-specialist audience.

Those who have already encountered Carter's earlier book *Mapping the Mind* (Weidenfeld & Nicolson, 1999) will have some idea of what to expect in style if not in content: a simplified but up-to-date summary of a difficult and controversial topic, embedded in glorious graphics and packaged in a lush coffee-table volume. Some readers may take issue with Carter's short-cuts: her tendency to report some speculation as fact, her lack of theoretical sophistication, and her tendency to gloss over some difficult areas of debate (What is a concept? How do language, perception and bodily awareness contribute towards a grounding of our experience?). But given the breadth of the material, substantial simplification is acceptable.

Exploring Consciousness does not have a strictly linear structure. There are numerous inserts in the main text, including brief essays from researchers and theoreticians, explanations of technical points, and descriptions of clinical phenomena. This innovative style worked well for *Mapping the Mind* but here it can occasionally distract the reader from Carter's attempts to develop her own thesis. The volume is almost crying out for electronic publication with a hyper-linked structure complete with animations and other multimedia aids.

As a conventional text, the organization of *Exploring Consciousness* accurately mirrors the fleeting focus and shifting perspectives that characterize our phenomenal awareness — and the many transient fashions of



Machinery of the mind: how the mind works, and what consciousness is, remain open to speculation.

consciousness research. In its present format, the book lends itself well to coffee-table or bathroom reading: it is possible to dip into the volume and feel satisfied by the refreshment. There are sufficient tasty morsels within to please the intellectual grazer. But the volume is less suited to those seeking a tightly structured statement of position or the sustained development of an argument.

The sheer scope of Carter's project is stunning and she handles her material with style and confidence, addressing a broad range of topics in the space of a few hundred pages. These range from an outline of historical and contemporary philosophical views on the relationship between brain and mind, through to discussions of the grounding problem, emotion and cognition, the nature of self, and the theory of mind — as

well as some weird stuff at the outer limits of scientific enquiry, such as altered states and religious experiences.

The book does contain some impressive pictures of brain scans, but it is far from being a revised edition of *Mapping the Mind*. In *Exploring Consciousness*, high-tech science plays a subsidiary role to the thorny theoretical issues. Carter uses illustrations from the imaging literature to explain some aspects of current thinking (such as the neural bases of emotion), but this book is more concerned with conceptual matters than with stunning pictures of the working brain.

Carter grapples with the unreliable nature of introspection, draws analogies between zombies and computational models of brain function, and considers the possibility that consciousness may be an epiphenomenon

without any survival value. I find the balance between technology and debate much more satisfactory here than in Carter's earlier book. Her approach and understanding seem to have matured, and localization of function is presented as one strand of evidence rather than as the ultimate solution to knotty abstract problems.

Consciousness has been a particularly fashionable area of enquiry for the past decade, and in this time some 30,000 articles have addressed themselves to the topic. But sadly, as this book makes clear, there is still no agreement either on the phenomena themselves or on the appropriate way to tackle them. Is consciousness the last gasp of cartesian dualism and a conceptual error, possibly reflecting our cultural predilection for a ghost in the machinery of the mind? Is it an adjective rather than a noun, describing the quality of experience rather than a distinct process or experiential context? Or were the behaviourists correct in putting the 'hard problem' of consciousness on one side in order to deal with the supposedly tractable problems of learning and responding? We are said to be waiting for a new Einstein to put consciousness on a secure theoretical footing, but the job is still vacant.

Carter is to be congratulated for her clear journalistic exposition of some key ideas and for avoiding pomposity and arrogance. The volume works, and although it is not a textbook, I shall certainly be recommending it to my second-year undergraduate students as a provocative source of ideas and an entrée into a difficult area. I shall also continue to enjoy the book myself. Good fare deserves digestion at leisure. ■

Rosaleen McCarthy is in the Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK.

Predicting extinction risk

Population Viability Analysis

edited by Steven R. Beissinger &

Dale R. McCullough

University of Chicago Press: 2002. 577 pp.

\$95 (hbk); \$35 (pbk)

Richard Frankham

Population viability analysis (PVA) is an important tool in conservation biology. It is the process of predicting the risk of extinction from the combined effects of the deterministic threats — such as habitat loss, overexploitation, pollution and species introductions — and stochastic threats, including demographic, environmental and genetic fluctuations and catastrophes. This is typically done using stochastic computer simulations. PVA is also used to compare

Still a round after all this time

Ammonites are among the most abundant yet beautifully preserved fossils on Earth. Many have tightly coiled spiral shells, although this *Scaphites nodosus* from the Cretaceous period is partly uncoiled. Although related to the modern-day nautilus, these cephalopods were actually more akin to the octopus, squid and cuttlefish. They died out at roughly the same time as the dinosaurs, after surviving for 300 million years before that. The biology of these extinct animals is the subject of *Ammonites* by Neale Monks and Phillip Palmer (Smithsonian Institution Press/The Natural History Museum, \$50 (hbk), \$24.95/£15.95 (pbk)).



alternative management options designed to help threatened species recover. The technique arose in the 1980s but is based on the accumulated knowledge from more than a century of research in demography, ecology and genetics. In some ways it is similar to weather forecasting and the modelling of economics and global climate.

Despite vigorous activity in this field, there have been few overviews of it. This long-awaited book, the authors of which are a veritable who's who of PVA, fills the gap. The book opens with an overview of PVA before moving on to consider the construction of PVA models, how to integrate theory and practice when using PVAs, and finally the future of PVA. The focus is primarily on animals throughout.

The book succeeds in reflecting the breadth of the field, the diversity of opinions about PVA, and its use in conservation biology. The contents are of variable quality, as is common in edited volumes. I found much to applaud and much to disagree with. The highlight for me was the chapter by Mark Shaffer and colleagues on PVA and conservation policy. It was incisive and thoughtful, and had a useful practical perspective, as befits a contribution from a founder of the discipline.

By contrast, the treatment of genetics throughout the book is sometimes dubious.

However, the chapter by Fred Allendorf and Nils Ryman provides a thorough, authoritative survey of the role of genetic factors in PVA. Several chapters refer to the problems of modelling genetic factors in PVA, but these apparent difficulties are largely illusory. VORTEX software can model inbreeding depression well for juvenile survival, and well-known functions can be used for other aspects of the life cycle. Furthermore, the chapter by Sue Haig and Jon Ballou describes work that encompasses inbreeding depression, as do other published papers. There are limited data for parameterizing inbreeding depression, but even that problem is slowly being addressed.

The book arose out of an international symposium in 1999, but many authors have updated their contributions to include more recent material. For example, our subsequent paper (*Nature* **404**, 385–387; 2000) on the predictive accuracy of PVA is referred to by several authors, although some describe its methods and contents incorrectly.

I was disappointed by the perspective provided by the book, despite there being many fine contributions. For example, the relationship of PVA with other related fields, which offer useful methods, analogies and insights, receives only cursory attention. The context of conservation biology as a crisis discipline, in which immediate decisions

must be made on the basis of inadequate data, seems to have been overlooked in many contributions. And several authors refer to alternatives to PVA but fail to point out that the only realistic alternative, human judgement and intuition, is inaccurate.

Deficiencies in data are a serious problem for PVAs of endangered species, as several authors point out. They discuss various remedies, including eliminating problems in data collection, separating the effects of sampling variation and intrinsic variation, and the use of bayesian methods to incorporate uncertainties about parameter values into predictions. However, the use of data from related species receives little attention. Such data may be particularly valuable for variances of input parameters. For example, environmental variation is similar across a range of herbivore taxa (J.-M. Gaillard *et al. Annu. Rev. Ecol. Syst.* **31**, 367–393; 2000). Furthermore, analyses of long-term data sets to estimate the frequency and severity of catastrophes and to ask whether they differ widely across taxa would be highly desirable.

The way to improve PVA is clear, but receives little attention here. It involves cycles of building models, making predictions, testing them, improving the models, and so on, as for all fields involving complex systems. The emphasis in PVA has been on building models, but the testing phase has only just begun. Much more testing is required if the field is to advance swiftly and rationally, as is happening with climate modelling.

Despite some limitations, this volume should serve as a major reference book on PVA for professional scientists, advanced undergraduates and graduate students in conservation biology. ■

Richard Frankham is in the Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia.

Shock results

Surprise, Uncertainty and Mental Structures

by Jerome Kagan

Harvard University Press: 2002. 259 pp.

\$29.95, £20.50

Peter Bryant

Developmental psychology — or at any rate the part that deals with children's intellectual abilities — is in an awkward position at the moment because it keeps on producing paradoxical and contradictory results. On one hand, a series of ingenious experiments has apparently shown that babies less than one year old possess some remarkable abilities. According to these experiments, infants can discriminate numbers, can add and subtract, and can distinguish cause and effect in both the physical and the social

worlds. On the other hand, a great deal of research on much older children, between the ages of three and about eight years, has apparently demonstrated that children in this age band often have serious difficulties with these same concepts. Their understanding of number appears fragile in such studies, their reasoning about arithmetical operations such as addition and subtraction is often erroneous, there are serious limitations to their grasp of cause and effect, and so on.

The contrast and apparent conflict between these two lines of research are stark and clear. Any suggestion that 6-month-old babies are much cleverer than children four and five years their senior would be highly implausible, and would be anathema to most developmental psychologists — their subject is dedicated to the proposition that intellectual skills improve throughout childhood. Another possible solution to this apparent impasse is to reappraise the work on infants. Are they as smart as this research suggests?

Jerome Kagan, who is a distinguished US developmental psychologist with an interest in children's emotional and intellectual development, has now provided such a reappraisal in his new book. He is well placed to do so because much of his research has been on children's expectations and their reactions to congruous and incongruous events. Nearly all of the work that apparently demonstrated various remarkable abilities

in infants used their reaction to novelty and their surprise at unexpected events as a way of measuring such abilities. If, for example, you add one object to another and then show the baby that the product of your actions is three objects, and if the baby is more surprised to see this as the result than she is to see two objects there, you can argue that the baby knows something about addition.

Kagan's discussion of this pervasive use of babies' reactions to novel and unexpected events comes in the second half of the book. It is based on ideas, which he expounds in the first part, about the ways in which children and adults form expectations and are surprised when these are not fulfilled. His reasoning is sophisticated and interesting, but his conclusions are easy to summarize. He is highly sceptical of the claims made about the presence of surprising abilities in infants, and argues that their achievements in these studies can be explained without ascribing huge intellectual acumen to them.

Kagan's reassessment of research on infants is the most important part of his new book. His arguments on the subject are controversial, of course, and should provoke a much-needed discussion. So the book will make a real contribution to the debate, and is definitely worth reading. ■

Peter Bryant is in the Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK.



Bright eyes? Babies' responses to surprising results may not be a good guide to their intelligence.

A chemical compass

Henry R. Bourne and Orion Weiner

Eukaryotic cells often devote much time and energy to business trips. Leukocytes crawl towards sites of infection, soil amoebae find other amoebae with which to form multicellular organisms, and budding-yeast cells find their mating partners. How do these cells know where to go? In each case, an internal cellular 'compass' interprets a shallow gradient of an external chemical attractant to orientate the cell's polarity in the correct direction; migration using this chemical compass is called chemotaxis. More than a billion years of eukaryotic evolution have conserved most components of the compass, yet we have just begun to scratch the surface of the network of positive and negative feedback loops that control gradient interpretation and cell polarity. Only now are we beginning to unravel this mechanism to reveal how a cell's ability to point in any direction allows it to navigate effectively.

The compass responsible for chemotaxis in eukaryotic cells differs from that of bacteria. Bacteria swim in a jerky trajectory, using a temporal strategy to interpret the gradient — they compare the concentration of attractant at one time with that of a moment before. In contrast, eukaryotic cells take advantage of their larger size to implement a spatial strategy — they compare attractant concentrations across the cell surface, point themselves towards the highest concentration, and move directly up the gradient. To do this, leukocytes such as circulating neutrophils in mammals amplify the shallow external gradient of the attractant by converting it to a steeper gradient of internal signals, resulting in a polarized cell with a protruding front that points up the external gradient. This polarity produces asymmetric attractant sensitivity — greater at the cell's protruding edge than at the trailing one. Consequently, a neutrophil confronted

with a 180° reversal in the attractant gradient does not simply transform its trailing edge into a leading one; instead, it follows its highly sensitive 'nose' and executes a U-turn.

We propose that asymmetric responsiveness to extracellular gradients involves two complementary mechanisms that make the cell's 'front' (pseudopod) distinct from its 'back'. Each mechanism depends on actin, a cytoskeletal protein, to produce a self-organizing system for creating morphological polarity. The first mechanism uses localized positive feedback to amplify responsiveness to the external signal at the front of the cell. The feedback loop promotes new assembly of actin polymers, which form the pseudopod, push it forwards, and facilitate accumulation of a lipid signal in the plasma membrane.

To confine positive feedback to the leading edge, a rapidly diffusing inhibitor acts to inhibit actin polymerization and membrane protrusion outside the pseudopod. This mechanism depends on assemblies of actin and the contractile protein myosin, located in the cell cortex at the rear and sides, which inhibit protrusion. Protrusion and contraction are controlled by different small enzymes that belong to the Rho family of proteins, which hydrolyse the nucleotide GTP (and are therefore called Rho GTPases). These two mechanisms cooperate to generate and stabilize an internal polarity that greatly exceeds the external gradient, allowing cells to detect and crawl persistently up shallow gradients of attractant and, in the case of neutrophils, to polarize in response to uniform attractant (usually in a randomly chosen direction).

In the positive feedback loop, the attractant causes an increase in the concentration of phosphatidylinositol-3,4,5-trisphosphate (PIP3), a membrane lipid that promotes Rho GTPase activation. In turn, the GTPases promote actin polymerization in a pseudopod at the leading edge, where the GTPases and new actin polymers together promote further PIP3 accumulation, amplifying the response to attractant at the front of the cell. The feedback loop is best documented in neutrophils, but PIP3 also accumulates at the front of the amoeba *Dictyostelium* and of fibroblast cells.

Inhibition at other parts of the cell confines the positive feedback loop to the leading edge, by promoting cortical actin-myosin assemblies and reducing membrane accumulation of PIP3. Disrupting the actin-myosin assemblies makes the back of a neutrophil as sensitive to attractant as the front — reversing the gradient causes the cell simply to reverse its direction, rather than performing a U-turn. Although necessary for long-term asymmetric sensitivity to chemoattractants, this contractile network is not absolutely required for

Cell polarity

Only now are we beginning to unravel the mechanisms behind a cell's ability to point in any direction and navigate effectively.

polarization itself — disruption of the network allows neutrophils and *Dictyostelium* to form PIP3-rich membrane protrusions towards a source of attractant, although the cells do form multiple pseudopodia at their sides, and crawl haltingly in the correct direction. In both cell types, membranes at the rear and sides fail to accumulate PIP3; in *Dictyostelium*, this resistance results from preferential association of a PIP3-degrading enzyme with these membranes. The actin-myosin network and mechanism of resistance to PIP3 accumulation cooperate, re-enforcing one another to interrupt the positive feedback loop that is necessary for protrusion.

During chemotaxis, gradient interpretation is intimately linked to the morphological polarity required for forward movement. The two essential cytoskeletal elements support diametrically opposite effects on PIP3 accumulation and attractant sensitivity, each limiting the proportion of the cell surface that is susceptible to regulation by the other. This conversation between opposite ends of the cell may be a self-organizing property of cytoskeletal systems that create cell polarity.

In one example of such a self-organizing system, round fragments of fish keratocytes respond to transient pressure on one side by polarizing and crawling away. Contractile actin-myosin complexes form on the pressured side, and actin polymerizes on the opposite side to form a protruding pseudopod; polarity and motility persist long after the stimulus disappears. In this case, as in chemotaxis, the front and back of the cell probably converse both mechanically (by asymmetric distribution of cortical tension) and biochemically (using Rho GTPases, PIP3 and other signals). A neutrophil decides correctly where to go because its 'nose' and its actin-myosin 'muscles' reciprocally regulate one another. These clever cells would scorn brains as an unnecessary luxury. ■

Henry Bourne is in the Departments of Medicine and of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143, USA. Orion Weiner is in the Department of Cell Biology, Harvard Medical School, Cambridge, Massachusetts 02115, USA.

FURTHER READING

Comer, F. I. & Parent, C. A. *Cell* **109**, 541–544 (2002). Verkhovsky, A. B. *et al. Curr. Biol.* **9**, 11–20 (1999). Weiner, O. D. *Curr. Opin. Cell Biol.* **14**, 196–202 (2002). Zigmond, S. H. *J. Cell Biol.* **75**, 606–616 (1977).

EYE OF SCIENCE/SPL



Stretching the point: the single-celled *Amoeba proteus* demonstrates its locomotor pseudopodia.

RNA as drug and antidote

Edward Tuddenham

Drugs that keep the blood circulation flowing help to prevent heart attacks and strokes caused by clots in the wrong places. A new anti-clotting tool is now revealed, along with its antidote.

The biggest cause of serious illness and premature death in Western societies today is the blockage of blood vessels by abnormal blood clots^{1–3}. On page 90 of this issue Rusconi and colleagues⁴ describe how they discovered a new drug that prevents blood clotting, and how they designed an antidote. The drug is an RNA ‘aptamer’, the sequence of which was found by iterative *in vitro* selection from a large combinatorial library of RNAs, screened for binding to a protein that promotes blood coagulation. Aptamers can be made of RNA or DNA, and they achieve their target-binding selectivity by virtue of the unique shapes they adopt through the formation of internal base pairs. Coagulation proteins are popular targets for both traditional and aptamer-based drugs, because they are available pure in soluble form, and their activity can be readily measured in assays of blood-clotting time or by their cleavage of small, colour-generating substrates⁵.

Coagulation proteins — or ‘factors’ — circulate in the bloodstream as inactive precursors, which sequentially activate each other after an initial trigger from tissue factor, exposed by damaged or diseased blood vessels (Fig. 1). An active coagulation factor is given the suffix ‘a’. So a complex of tissue factor and factor VIIa activates factor IX (ref. 6) — the target of Rusconi and colleagues’ aptamer⁴. Factor IXa, together with factor VIIIa, then activates factor X. Next, factor Xa, with help from factor Va, activates factor II (also called prothrombin), producing thrombin. Finally, thrombin converts soluble fibrinogen to fibrin, the protein that polymerizes to form a solid blood clot. Thrombin also activates and promotes the aggregation of platelets, which are blood cells that help to form a plug in the wall of the wounded vessel.

Normally, this process represents a vital first line of defence against blood loss. But a clot or platelet plug in the wrong place can cause a heart attack or stroke, by preventing blood flow and therefore oxygen supply to a particular area of tissue — hence the importance of drugs that inhibit blood clotting. Drugs that block the production of thrombin are termed anticoagulants, whereas those that inhibit platelet function are called antiplatelet agents. Remarkably, until quite recently there was only one class of antiplatelet agent, aspirin and its relatives, and only two classes of anticoagulant,

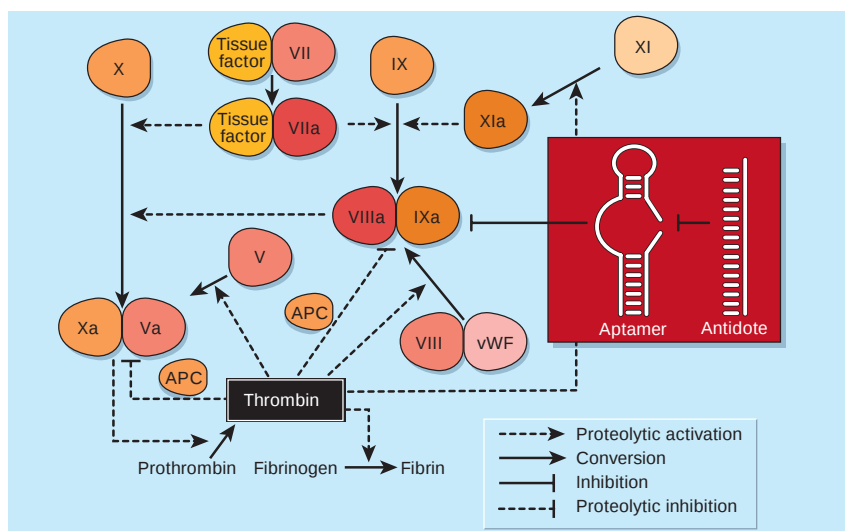


Figure 1 How blood clots. From the top: clotting is triggered when factor VII, circulating in the blood, encounters tissue factor, exposed by blood-vessel damage or disease. The complex of tissue factor and factor VIIa (activated factor VII) activates factor IX, which in turn, with factor VIIIa, activates factor X. With factor Va, factor Xa converts prothrombin to thrombin; thrombin then converts fibrinogen to fibrin; and fibrin forms blood clots. There are various checks and balances — for example, factors VIII and V are back-activated by thrombin and can be degraded by activated protein C (APC), produced by thrombin in the presence of thrombomodulin (found on intact blood-vessel walls). Also, the complex of tissue factor and factor VIIa can activate factor X both indirectly (via factor IX) and directly: the direct pathway is rapidly shut off by tissue-factor-pathway inhibitor, whereas factor IX can be further activated by factor XI (top right), which is itself activated by thrombin. Many proteins and reactions have been omitted for clarity. Rusconi *et al.*⁴ have developed an RNA-based drug (an aptamer) that inhibits factor IXa, and its antidote. vWF, von Willebrand factor.

heparins and antagonists of vitamin K, in widespread clinical use. Even more surprisingly, these were discovered (or introduced to the clinic) as long ago as 1900, 1916 and 1940, respectively. So the anti-clotting armamentarium remained sparsely stocked for most of the past century. Some innovative approaches have, however, started to fill it.

Beginning with leeches (in medicinal use for at least 1,000 years), researchers have screened a series of blood-feeding invertebrates and even vampire bats for the presence of anticoagulants in their digestive tracts — obviously a requirement for successfully eating blood. Hirudin is a thrombin inhibitor in leech saliva, and genetically engineered hirudin, as well as a second-generation derivative⁷, is now in clinical use for a few conditions. An anticoagulant peptide from the soft tick is a highly selective inhibitor of activated factor X, and a peptide from hookworms inhibits factor VII. The giant Amazon leech has a unique platelet-

disaggregating enzyme in its saliva⁸, and vampire bats keep the blood flowing with a powerful clot-dissolving enzyme. None of these peptides or enzymes has reached the clinic yet. But several small synthetic molecules that inhibit thrombin or factor Xa directly are in clinical trials⁹. Moreover, a mouse antibody, modified to resemble human antibodies, that blocks a fibrinogen receptor on the platelet membrane is in widespread clinical use¹⁰.

All of these newer anti-clotting weapons have a drawback, however: they have no rapidly acting antidote. Consequently, overdose may result in serious bleeding, with no effective treatment. The older drugs — heparin, warfarin and aspirin — are not wholly immune to this criticism either, although protamine, a protein from fish sperm, can be used to neutralize heparin by virtue of its positive charge (heparin is negatively charged). An overdose of warfarin severely reduces levels of all the vitamin-K-

dependent coagulation factors, and can be reversed in an emergency by plasma infusion — but this carries a risk of virus transmission. The use of aspirin is associated with a high risk of stomach bleeding; it also leads to bleeding in the brain in a few long-term users. Aspirin is an irreversible inhibitor of a platelet enzyme, cyclo-oxygenase, and no antidote is available, or indeed possible. Ticlopidine, which also inhibits platelets, similarly suffers the ‘no antidote’ problem.

So Rusconi *et al.*⁴ wanted to see whether they could design both an anticoagulant and its matching antidote by taking advantage of the properties of RNA. The authors’ choice of factor IX as their target is significant: previous animal studies¹¹ involving the infusion of factor IXa that had an inhibited active site revealed a favourable therapeutic margin compared with heparin in preventing experimentally induced clotting. (‘Therapeutic margin’ refers to the dosage difference between preventing clotting and inducing blood-vessel failure through overt bleeding.) Until now, however, searches of natural products and screens of chemical libraries have not revealed a selective inhibitor of factor IXa.

Rusconi *et al.* started with eight rounds of selection for factor-IXa-binding aptamers from a 10¹⁴-fold combinatorial library of 80-nucleotide RNAs. They thereby generated an aptamer that is 5,000 times more specific for factor IXa than for the structurally related factors VIIa, Xa, XIa and activated protein C. This last point is important, because a protein C inhibitor would promote blood clotting (see Fig. 1). By comparing the effective sequences, the authors generated a truncated form to which polyethylene glycol could be attached without affecting the aptamer’s affinity for factor IXa or its anticoagulant activity. (Polyethylene glycol increases the time that RNA aptamers can spend in the blood circulation.) Finally, although Rusconi *et al.* conducted all their studies *in vitro*, they did detect a significantly longer clotting time, compared with controls, after adding their aptamer to blood plasma.

What of the antidote? Aptamers achieve selectivity by adopting unique three-dimensional folds, as a result of internal Watson–Crick base pairing. To disrupt these folds it is simply necessary to design a nucleotide sequence that will compete effectively for the critical base pairs. Rusconi *et al.* have done this convincingly, using a complementary sequence to reverse the anticoagulant effects of the selected aptamer completely and rapidly.

How soon can we expect to see this class of drug and its antidote in clinical use? First we have to wait for *in vivo* studies in animals, which are now in progress. There is no reason to suppose that the aptamer and antidote would be toxic, or that they would affect clotting *in vivo* differently from that *in vitro*.

Nonetheless, it is impossible to predict whether other enzymes or physiological systems will be adversely affected until the animal — and, ultimately, human — studies have been carried out.

It has been proposed that life on Earth had its origins in RNA, which had both a passive role as an information store and an active role as an enzyme that propagates this information. Later, these tasks became split between DNA and proteins. But perhaps we have now come full circle: the work of Rusconi *et al.*⁴ shows how one RNA molecule can have an active role as a drug, and can store the information required to make its own perfect antidote. ■

Edward Tuddenham is in the Faculty of Medicine, Imperial College of Science, Technology and

Medicine, Haemostasis Research, MRC Clinical Sciences Centre, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK.
e-mail: edward.tuddenham@csc.mrc.ac.uk

1. Hirsh, J. & Lee, A. Y. *Blood* **99**, 3102–3110 (2002).
2. Turpie, A. G. & Antman, E. M. *Arch. Intern. Med.* **161**, 1484–1490 (2001).
3. Mannucci, P. M. & Poller, L. *Br. J. Haematol.* **114**, 258–270 (2001).
4. Rusconi, C. P. *et al.* *Nature* **419**, 90–94 (2002).
5. Bock, L. C. *et al.* *Nature* **355**, 564–566 (1992).
6. Eigenbrot, C. & Kirchhofer, D. *Trends Cardiovasc. Med.* **12**, 19–26 (2002).
7. Carswell, C. I. & Plosker, G. L. *Drugs* **62**, 841–870 (2002).
8. Sawyer, R. T., Jones, C. P. & Munro, R. *Blood Coag. Fibrinol.* **2**, 153–159 (1991).
9. Pauls, H. W. & Ewing, W. R. *Curr. Top. Med. Chem.* **1**, 83–100 (2001).
10. Harnick, D. J. & Vorchheimer, D. A. *Curr. Cardiol. Rep.* **3**, 355–361 (2001).
11. Himber, J. *et al.* *Thromb. Haemost.* **85**, 475–481 (2001).

High-energy physics

The matter with antimatter

Michael Peskin

The Universe is made of matter, not antimatter, and ‘CP violation’ in particle decays could be the reason. Results from experiments measuring this effect at last confirm the predictions of a 30-year-old theory.

Every elementary particle has an antiparticle, a counterpart with precisely the same mass and the opposite electric charge. It would seem natural that all of the interactions of antiparticles are just the opposite of those of particles. But in 1964, in a Nobel-prize-winning experiment, Cronin and Fitch showed that this is not so¹. Their measurement showed up a tiny difference, one part in a million of the strength of the weak interaction. New measurements from the BaBar experiment² at Stanford, USA, and from the Belle experiment³ at Tsukuba, Japan, seem finally to have established a definite origin for this small difference in the behaviour of particles and antiparticles. The new data are a triumph for a theory proposed 30 years ago. But they are also a disappointment. The results so far do not help us understand the most obvious matter–antimatter asymmetry in nature — the fact that we see matter, but no antimatter, in the Universe at large.

The asymmetry in question is a basic one. It is well established that the laws of physics are the same irrespective of the position, time or orientation of the observer. From this statement, it would seem obvious that the laws of physics would also be the same when reflected in a mirror. But in 1956, following the suggestion of Lee and Yang, it was found that the weak interactions that mediate radioactive decay are completely mirror-asymmetric: nuclear β -decay preferentially produces particles that spin in the left-handed sense and antiparticles that spin in the right-handed sense. The interaction, however, seemed to respect the combined

operation of mirror reflection (which turns a left-handed spin into a right-handed one) and interchange of particle and antiparticle. Physicists refer to these two operations as P (‘parity’) and C (‘charge conjugation’), so the combined operation is called CP.

The Cronin–Fitch experiment, however, showed that the CP symmetry could be violated in specific particle decays. But it was not clear how to develop new equations to include this effect. CP violation arises only when particle masses are generated, and the origin of mass in particle physics is even now a mysterious issue, one that the elusive Higgs boson might solve.

Conversely, CP violation, not C or P violation alone, is required if an unstable heavy particle decays asymmetrically into matter and antimatter. Just after the Cronin–Fitch result, Andrei Sakharov suggested that CP violation could make particle decays in the early Universe asymmetric and thus could create an excess of matter from a hot Universe containing equal amounts of matter and antimatter⁴. The idea that the present Universe contains only matter as the result of a CP asymmetry in the fundamental laws is very attractive, but to evaluate it we need to know the origin of CP violation. CP violation arises in expressions for particle masses. In quantum mechanics, a particle with mass is a quantum state, which may be a mixture of some more basic states. CP violation results from that mixing, when the mixing amplitudes are complex numbers with different phases. But which specific particles give rise to the CP violation that we observe?

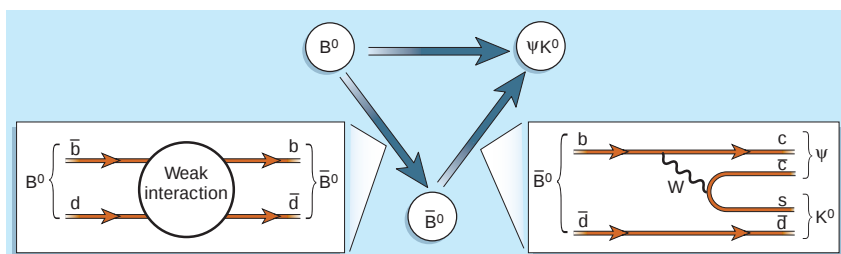


Figure 1 Decay of a B meson. The B^0 meson (made of bottom and down quarks) can decay into two other mesons, known as ψ and K^0 (made of charm quarks and strange and down quarks respectively). The decay can occur directly, or the B^0 can convert to its antiparticle, the $\bar{B}^0$ meson, through a weak interaction between its constituent quarks (left); the antiparticle can also decay to ψK^0 (right) through another weak interaction involving the W boson. Because there are two pathways to the final state, quantum interference arises between them. The BaBar² and Belle³ experiments exploit this as a means of measuring CP violation.

In 1972, when the model of weak interactions was still speculation and even quarks (the elementary particles that are the constituents of particles such as the proton and the neutron) were viewed with scepticism, Makoto Kobayashi and Toshihide Maskawa put forward the idea⁵ that CP violation originates in the mass mixing of heavy quarks. Quarks were expected to have mass mixing; Cabibbo and others had shown this theoretically^{6,7}. At that time, though, only three species of quarks were known (called up, down and strange), and all three had small masses. Kobayashi and Maskawa required more quarks, six in all, to build a theory with intrinsic phase differences in the mixing terms. It was a wild idea. But, wild or not, these extra, heavy quarks (charm, bottom and top) have since turned up in experiments — the sixth, top, quark only in 1995.

If heavy quarks are needed for CP violation, it is logical that the decays of heavy quarks would show much larger CP asymmetries. Bigi, Carter and Sanda^{8,9} proposed looking in the decay of a B^0 meson (a particle that contains a heavy bottom, or b, quark together with a light quark) and specifically in its decay to two other mesons, called ψ and K^0 . This decay can occur by two different paths (Fig. 1): either the B^0 can convert directly to ψK^0 , or it can convert to its antiparticle $\bar{B}^0$, which then turns into ψK^0 . In quantum mechanics, when a reaction can occur by two different paths, interference results. In this well-chosen reaction, the relative phase of the interfering terms is precisely the CP-violating phase of the Cabibbo–Kobayashi–Maskawa (CKM) model. If CP violation is fundamentally a large effect, the phase should be large.

Experimentally, measuring this phase is challenging. This particular decay can be observed in only 1 in 100,000 B decays, requiring huge numbers of high-energy collisions and unprecedented accelerator performance. New accelerators were built in Japan and the United States to meet these needs and have been operating for three years. They now hold nearly all the performance records for

colliding-beam particle accelerators, and their corresponding detectors, Belle and BaBar, have collected a total of 300 million B decays.

At this summer's International Conference of High Energy Physics in Amsterdam, the BaBar and Belle experimental groups presented the results on the CKM phase that they have derived from these data^{2,3}. The two results are precise and consistent. Remarkably, the value of the CKM phase they measure is exactly that needed to explain the magnitude of CP violation seen in the Cronin–Fitch experiment nearly 40 years ago. In the next few years, these measurements of B decay will be sharpened and new measurements from BaBar and Belle, and from the experiments at the Fermilab Tevatron, will come into play.

There is much at stake, for, despite many attempts, no one has been able to use the CKM model of CP violation to create the matter excess of the Universe through Sakharov's mechanism. Only a small asymmetry is needed in the early Universe, as today we have only one leftover proton for 10^9 photons. But simple calculations in the CKM model give a prediction of one proton to 10^{18} photons, and no cosmological model has been found that improves this by more than a few orders of magnitude. The problem is that CP violation in the CKM model requires not only the heavy-quark but also the light-quark masses, and these latter terms are unimportant at the high temperatures of the early Universe¹⁰. Models of CP violation that involve heavy objects only — for example, models with additional CP-violating phases in the mass mixing of Higgs bosons or other undiscovered particles — can readily account for the observed excess of protons.

Unless the new particles are extremely heavy, their CP effects should also show up in B decays. If these particles are present, the precise B-decay experiments of the next few years should show anomalies: perhaps qualitative differences from the CKM predictions, perhaps 10–20% discrepancies that the CKM model cannot account for. At present, no one knows what the result will be. Later in the decade, millions of top quarks and other



100 YEARS AGO

A careful experimental inquiry regarding the nutritive value of alcohol has recently been carried out in the chemical laboratory of Wesleyan University... The main question studied is the value of alcohol as a fuel in the human body and its comparison in this respect with sugar, starch, fats and other nutrients of ordinary food materials. Collaterally, the question of the effect of alcohol upon the proportions of nutrients digested from the food with which it was taken has also been examined. Metabolic experiments on an elaborate scale have been instituted with the view of investigating the problem, and no expense has been spared to obtain complete and accurate results... The results of the inquiry indicate that more than 98 per cent. of the ingested alcohol was oxidised in the body and that the potential energy of the alcohol was transformed into kinetic energy as completely as that of the ordinary nutrients... The conclusion is drawn that so far as the utilisation of the total energy of the diet is concerned, there is a slight advantage in favour of the non-alcoholic diet, especially when the body is subjected to hard muscular exertion, but the difference is so small as to lie almost within the limits of experimental error.

From *Nature* 4 September 1902.

50 YEARS AGO

An investigation into the interests of about 391 people (241 males, 150 females) who are taking science subjects in adult education classes showed that the females were more extreme in their preferences and dislikes than the males... Males were most interested in studying future advances and new discoveries and theories in science and, next to these, medicine, health and the industrial applications of science. Medicine and health were the most popular subjects for females, and almost as popular were pure biology and psychology. Neither males nor females took much interest in aeronautics or in the public direction and use of science; even less interest was shown in the work of scientific institutions and the results of science surveys and commissions. This may indicate that the 'ordinary' man is not fully conscious of the impacts of science on communal life or, perhaps, is little interested in the development of society as distinct from the individual.

From *Nature* 6 September 1952.

exotic particles will be produced at CERN's Large Hadron Collider in Geneva, Switzerland, offering an opportunity to search directly for CP violation associated with these heavy states.

For the moment, Kobayashi and Maskawa appear triumphant. Their 1972 model of heavy quarks seems to explain all of the CP violation seen today in particle-physics experiments. If it cannot also explain the Universe, that is a task for stranger particles still waiting to be discovered.

Michael Peskin is in the Theoretical Physics Group, Stanford Linear Accelerator Center, Stanford

University, Stanford, California 94309, USA.

e-mail: mpeskin@slac.stanford.edu

- Christenson, J. H., Cronin, J. W., Fitch, V. L. & Turlay, R. *Phys. Rev. Lett.* **13**, 138–140 (1964).
- Aubert, B. *et al.* (BaBar collaboration) Preprint hep-ex/0207042 (2002); <http://arXiv.org>
- Abe, K. *et al.* (Belle collaboration) Preprint hep-ex/0207098 (2002); <http://arXiv.org>
- Sakharov, A. D. *JETP Lett.* **5**, 24–27 (1967).
- Kobayashi, M. & Maskawa, T. *Prog. Theor. Phys.* **49**, 652–657 (1973).
- Cabibbo, N. *Phys. Rev. Lett.* **10**, 531–533 (1963).
- Glashow, S. L., Iliopoulos, J. & Maiani, L. *Phys. Rev. D* **2**, 1285–1292 (1970).
- Bigi, I. I. & Sanda, A. I. *Nucl. Phys. B* **193**, 85–108 (1981).
- Carter, A. B. & Sanda, A. I. *Phys. Rev. D* **23**, 1567–1579 (1981).
- Huet, P. & Sather, E. *Phys. Rev. D* **51**, 379–394 (1995).

Cell biology

Spinning actin to divide

Shuh Narumiya and Issei Mabuchi

When our cells divide, they are cut down the middle by a tightening belt of proteins. New work reveals that the protein filaments in this belt are made from scratch every time.

The ability of cells to multiply lies at the heart of many biological processes. In multicellular organisms such as ourselves, cell proliferation is essential for growth and development, and to replace cells spent by daily wear and tear. For single-celled species such as yeasts, proliferation is crucial because it is how these organisms reproduce. In order to proliferate, a cell has to duplicate its contents and then divide physically into two, distributing the duplicated contents evenly between the two new cells. This process of cell division — called cytokinesis — is carried out rather as if a thread encircling a boiled egg was gradually pulled tighter to constrict the egg and cut it across the middle. On page 82 of this issue, Pelham and Chang¹ provide a clear picture of how the cell spins such a thread.

The thread encircling a dividing cell is a belt of proteins called the contractile ring (Fig. 1) and — unlike the thread around the boiled egg — is constructed within the cell, just beneath the cell membrane. The ring is composed of actin filaments and myosin proteins, well-known components of muscle. It is thought to constrict when oppositely orientated actin filaments slide over each other with the help of myosin, just as in contracting muscle.

Actin filaments are long helical polymers of globular actin monomers. The conventional view of how the contractile ring forms is that ready-made actin filaments in the cell periphery are recruited to the site of cytokinesis during or after segregation of the genetic material (mitosis), and are assembled into the ring by myosin or by actin-bundling proteins. But last year it was shown in dividing frog eggs that actin monomers are rapidly incorporated into actin filaments in the

contractile ring as it is being constructed, implying that formation of new actin filaments is required². So what does happen? Pelham and Chang's detailed analysis¹ finds that the actin filaments are made from scratch and continuously assemble and disassemble. The authors also identify proteins that control this process.

All actin filaments form in two steps (Fig. 1). In the first — nucleation — three or four actin monomers assemble into a cluster, which acts as a seed. In the second step, called polymerization or elongation, monomers are successively added to the seeds, allowing rapid growth of the filament. Nucleation is the rate-limiting step, but is speeded up by a multiprotein complex named the Arp2/3 complex³.

Pelham and Chang started by analysing the involvement of *de novo* actin-filament formation in the construction of

the contractile ring in the fission yeast *Schizosaccharomyces pombe*, which divides in a similar way to mammalian cells. They used an *in vitro* test of ring assembly, which involves permeabilizing the dividing yeast cells, and incubating them with fluorescently labelled actin monomers. If actin nucleation and polymerization occur, the fluorescent monomers will be incorporated into the ring. Indeed, the authors found that a fluorescent ring was produced in these cells, and that ring formation was sensitive to an inhibitor of actin polymerization. In contrast, a drug that caps the rapidly growing ends of existing actin filaments did not affect the rate of actin incorporation into the ring. So the results imply that actin filaments are formed from scratch and *in situ* in the contractile ring.

Pelham and Chang then looked at the contribution of the Arp2/3 complex to actin dynamics in the permeabilized-cell system. They found, first, that mutant cells lacking Arp3 did not show actin incorporation. Second, antibodies that bind to Arp3 interfere with incorporation of actin into the contractile ring. So the Arp2/3 complex is necessary for formation of contractile-ring actin filaments, presumably at the nucleation step.

The authors also reveal that another means of controlling filament formation is involved in ring construction *in vivo* in *S. pombe*. This regulatory mechanism consists of two proteins, Cdc12 and Cdc3. Cdc12 is a member of the formin family of proteins, which are found in organisms from yeasts to mammals and include Bni1 in budding yeast, Diaphanous in fruitflies and mDia in mammals⁴. Cdc3 is an actin-monomer-binding protein, also known as profilin. Formins related to Diaphanous bind profilin, and both are essential in cytokinesis^{4,5}.

Interestingly, the Arp2/3 complex and the formins have been shown to induce the formation of distinct actin structures: patches and cables, respectively, in budding yeast^{6,7}; and lamellipodia (sheet-like cellular

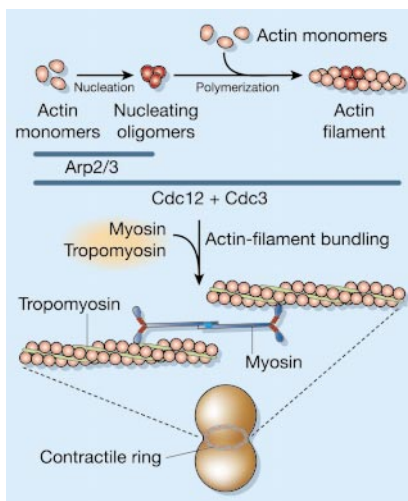


Figure 1 Spinning out the contractile ring. This ring, shown at the bottom, constricts to divide a proliferating cell into two. It is composed of antiparallel actin filaments, cross-linked by myosin molecules. Pelham and Chang¹ have found that the filaments are formed from scratch at the site of division, and continuously assemble and disassemble. The process of filament formation is shown from the top. First, actin monomers assemble into trimers ('nucleation'); trimers act as a seed to which further monomers are added ('polymerization'). Pelham and Chang find that the Arp2/3 complex and the Cdc12 and Cdc3 proteins are needed for nucleation or polymerization. The filaments are then organized into the contractile ring.

extensions) and stress fibres, respectively, in mammalian cells^{3,8}. More intriguingly, the formin Bni1 can nucleate actin polymerization independently of the Arp2/3 complex *in vitro*, and the presence of profilin markedly accelerates subsequent polymerization^{9,10}. It seems, then, that two mechanisms for actin nucleation and polymerization — one involving the Arp2/3 complex, and the other requiring formins and profilins — are recruited to, and work at, the site of cell division.

Finally, Pelham and Chang studied the stability of the contractile ring, and found that it is highly dynamic, with its components exchanging every minute. To show this, they used the technique of 'fluorescence recovery after photobleaching'. They tagged the protein tropomyosin or a subunit of myosin — both of which are actin-binding components of the contractile ring — with green fluorescent protein. When they bleached the fluorescence with laser light, they found that it recovered with a half-time of less than 30 seconds, indicating that the ring is broken down and reconstructed with new fluorescent components within this time period. Similar dynamics of myosin in the contractile ring have been reported previously¹¹, but the rate of exchange determined by Pelham and Chang is much faster. This requirement for a continuous supply of filament-binding proteins implies that not only actin polymerization, but also actin depolymerization, occurs in the ring, and explains the previously puzzling requirement for the actin-depolymerizing factor ADF/cofilin¹².

So Pelham and Chang¹ have shown that the contractile ring undergoes continuous remodelling to carry out its functions. Given the previous observations of dividing frog eggs², it seems that this mechanism is not limited to fission yeast, but applies more generally. What are the implications of these results? One of the authors' main findings is the need for the formin Cdc12 in ring construction. Although not explicitly shown for Cdc12, Diaphanous-related formins are targets of the Rho protein^{4,5}, which works as a molecular switch in various cell processes and has been suggested to regulate cytokinesis¹³. So this new study once again highlights the role of Rho in cell division. Studies of the functions and dynamics of Rho in cytokinesis might resolve several issues, such as how the plane of division is determined, and how the timing of cleavage is regulated. ■

Shuh Narumiya is in the Department of Pharmacology, Kyoto University Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan.

e-mail: snaru@mfour.med.kyoto-u.ac.jp

Issei Mabuchi is in the Department of Biology, School of Arts and Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan.

e-mail: mabuchi@ims.u-tokyo.ac.jp

1. Pelham, R. J. & Chang, F. *Nature* **419**, 82–86 (2002).

2. Noguchi, T. & Mabuchi, I. *J. Cell Sci.* **114**, 401–412 (2001).

3. Higgs, H. N. & Pollard, T. D. *Annu. Rev. Biochem.* **70**, 649–676 (2001).

4. Wasserman, S. *Trends Cell Biol.* **8**, 111–115 (1998).

5. Watanabe, N. *et al. EMBO J.* **16**, 3044–3056 (1997).

6. Evangelista, M. *et al. Nature Cell Biol.* **4**, 32–41 (2002).

7. Sagot, I., Klee, S. K. & Pellman, D. *Nature Cell Biol.* **4**, 42–50 (2002).

8. Watanabe, N. *et al. Nature Cell Biol.* **1**, 136–143 (1999).

9. Pruyne, D. *et al. Science* **297**, 612–615 (2002).

10. Sagot, I. *et al. Nature Cell Biol.* **4**, 626–631 (2002); advance online publication, 22 July 2002 (doi:10.1038/ncb834).

11. Wong, K. C. Y. *et al. Curr. Biol.* **12**, 724–729 (2002).

12. Abe, H. *et al. J. Cell Biol.* **132**, 871–885 (1996).

13. Mabuchi, I. *et al. Zygote* **1**, 325–331 (1993).

Materials science

Edge effects

Michael O'Keeffe

How do crystal structures terminate at 'flat' surfaces? New developments in electron crystallography mean that the detailed atomic structure of surfaces in complex crystals can be determined — with surprising results.

Anyone who has made a model of a crystal structure, or even just a drawing of one, must surely be struck by the fact that there is rarely, if ever, a satisfactory way of terminating the structure. Almost invariably, atoms with missing bonds are left dangling at the surface, and the chemist, at least, will immediately conclude that there must be some rearrangement at the surface to produce a more satisfactory bonding configuration. Remarkably, except for a handful of elemental and simple binary materials, until now we have had no knowledge of the real structure of most crystal surfaces, despite their enormous importance. But on page 55 of this issue, Erdman *et al.*¹ show how direct-methods electron crystallography coupled with *ab initio* electronic structure calculations can be used to determine a complex surface structure of strontium titanate, SrTiO₃. The results are surprising (although, with hindsight, perhaps to be expected) and they suggest that a new era of surface crystallography is beginning.

Surface structures are important in understanding processes such as catalysis, in which chemical reactions occur at the interface between a solid and a liquid or gas phase. These reactions are ubiquitous in the chemical industry, as well as at electrodes in fuel cells and in the dissolution (corrosion) of solids. Crucial to understanding the mechanisms of all these processes is a knowledge of the structure of the interface between the solid and the fluid medium. This last point leads to the observation that a surface is an interface between two phases, and in general the surface of a solid might have a different structure when it is the interface between a crystal and, for example, a vacuum or water or reacting gases.

Oxides are very important in this respect, and Erdman *et al.*¹ have chosen a particularly appropriate material for their study: SrTiO₃ is an example of a perovskite material, perhaps the most studied class of oxide. The parent crystal structure is relatively simple and consists of a cubic framework of corner-connected TiO₆ octahedra, with Sr atoms in the cavities of the structure (Fig. 1). Related

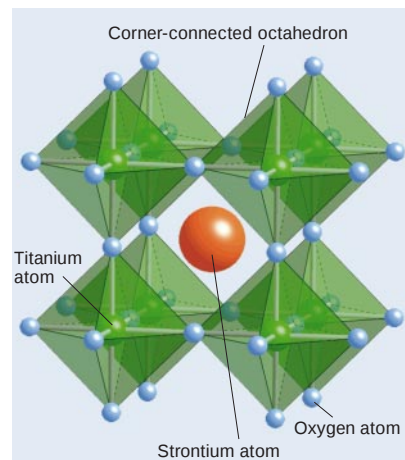


Figure 1 The crystal structure of SrTiO₃. Each titanium atom is bonded to six oxygen atoms, forming an octahedral structure that is connected at its corners to its neighbours. Combining calculations and crystallographic images, Erdman *et al.*¹ show that, at a surface in the horizontal plane of this crystal, the structure is reorganized, terminating in a double Ti–O layer with connections between polyhedral edges as well as corners.

members of its large family exhibit a remarkable range of compositions and structures². Perovskites include ferroelectric materials such as BaTiO₃ (the prototypical oxide ferroelectric) and related high-dielectric-constant materials. Other transition-metal perovskites exhibit exceptional electronic properties such as colossal magnetoresistance and (in oxygen-deficient copper oxides) high-temperature superconductivity. SrTiO₃ is often used as a substrate for epitaxial growth of such materials, and this makes its surface structure particularly interesting.

Because of their much stronger interaction with matter, electrons (rather than X-rays or neutrons) are the preferred probe for the examination of surface structure. An early use of electron microscopy was the determination of simple metallic surface structures from end-on ('profile') images³. Shortly thereafter, the structure of the famous

Si 7×7 surface was determined⁴ from transmission electron diffraction combined with scanning tunnelling microscopy data.

In determining the structure of crystals by X-ray diffraction, a major problem is that one knows only the intensities and not the phases of the diffracted beams. So direct inversion of the diffraction pattern by Fourier analysis to recover the electron density (which scattered the X-rays) is not possible. However, by recognizing that the electron density is not an arbitrary function but has special properties (for example, it is always positive and, as atomic electron density peaks strongly near the nucleus, is close to zero between atoms), relationships between the phases of diffracted beams can be established. These can then be factored into 'direct methods' of crystallographic analysis, which have made routine the determination of structures of moderately complex crystals (some hundreds of atoms in the unit cell). Electron crystallography is in principle similar — the scattering now occurs by the electrostatic potential of the atom, but this is simply related to the charge density by Poisson's equation — and direct methods have been used here too⁵.

It is not straightforward to apply direct methods to the solution of surface structures by electron crystallography, but Erdman *et al.*¹ show that direct-methods techniques developed by Marks and colleagues⁶ lead directly to a projection of the surface structure. They then use *ab initio* electronic-structure calculations to refine the coordinates normal to the surface to obtain a full three-dimensional structure.

In the SrTiO₃ structure (Fig. 1), there are alternating SrO and TiO₂ layers normal to the cubic axes. So one might expect a surface in that plane to terminate either with a TiO₂ layer or a SrO layer. In fact, in the observed structure¹ the crystal terminates with two TiO₂ layers, and with substantial reorganization in the outermost layer. The origin of the reorganization can be seen as follows. In the bulk structure each O atom is bonded to two Ti atoms. If this pattern were to persist in a surface of composition TiO₂, each Ti atom would have to be 4-coordinated (that is, forming four bonds to oxygen) in contrast to the octahedral 6-coordination in the bulk. The surface reconstruction is such that coordination polyhedra share edges (rather than just vertices as in the bulk) so that the Ti atoms in the surface layer are 5-coordinated.

Such behaviour might have been expected from crystal chemistry. In the tungsten oxide WO₃, for example, the structure is the same as the TiO₃ framework of SrTiO₃. WO₃ is easily reduced to phases WO_{3-x} in which, instead of having oxygen vacancies, the crystal reorganizes to eliminate the vacancies and maintain the octahedral coordination of the metal by edge-sharing between the octahedra. Similar behaviour is observed in the

fantastically rich and highly organized structure of the many other oxides of the early transition metals (such as Ti, Nb, Mo and W)⁷.

It should be emphasized that the work under discussion is the beginning of a long and surely very interesting story. Even for SrTiO₃ there is strong evidence⁸ for other reconstructions of the surface; in addition, as well as the perovskites A²⁺B⁴⁺O₃ (such as SrTiO₃), there are other families A³⁺B⁵⁺O₃ (such as NaNbO₃) and A³⁺B³⁺O₃ (such as LaFeO₃). And then, of course, there are many thousands of other oxide structures. ■

Michael O'Keeffe is in the Department of Chemistry

and Biochemistry, Arizona State University, Tempe, Arizona 85287, USA.

e-mail: mokeeffe@asu.edu

1. Erdman, N. *et al.* *Nature* **419**, 55–58 (2002).
2. Mitchell, R. H. *Perovskites: Modern and Ancient* (Almaz, Thunder Bay, Ontario, 2002).
3. Marks, L. D. *Phys. Rev. Lett.* **51**, 1000–1002 (1983).
4. Takayanagi, K., Tanishiro, Y., Takahashi, S. & Takahashi, M. *Surf. Sci.* **164**, 367–392 (1985).
5. Dorset, D. L. *Structural Electron Crystallography* (Plenum, New York, 1995).
6. Marks, L. D., Erdman, N. & Subramanian, A. *J. Phys. Condens. Matter* **13**, 10677–10688 (2001).
7. Hyde, B. G. & Andersson, S. *Inorganic Crystal Structures* (Wiley, New York, 1989).
8. Jiang, Q. D. & Zegenhagen, J. *Surf. Sci.* **425**, 343–354 (1999).

Immunology

The roots of antibody diversity

Patricia J. Gearhart

When faced with foreign molecules our antibodies mutate, allowing them to bind to the intruders more strongly. In a story full of surprises, it looks as though the mechanism of mutation has finally been revealed.

One of the tactics our immune system uses to fight off intruders is the production of protective antibody proteins, which recognize and neutralize foreign molecules. Antibodies are secreted by the B cells of the immune system, and are unique among our proteins in their unlimited potential for diversity. Half of each antibody molecule is encoded by the so-called variable genes, and when a B cell recognizes and responds to a foreign molecule, these variable genes undergo mutation at a tremendously high frequency.

In a stunning set of papers, culminating in that on page 43 of this issue¹, Neuberger and colleagues seem to have discovered the elusive mechanism behind this 'hypermutation' — and the answer is totally unexpected. First they showed that the activation-induced cytosine deaminase (AID) protein, which is needed for hypermutation², leads to DNA mutations when expressed in bacteria³. More mutations of the cytosine 'letters' in DNA occur in bacteria that lack the enzyme uracil glycosylase, hinting that AID converts cytosines to uracils. (Normally, these erroneous uracils would be removed by the glycosylase before they caused a problem.) These findings are now followed⁴ by a convincing demonstration that the generation of uracil from cytosine in B-cell DNA sparks hypermutation, which occurs as the cells try to fix the errors.

After vaccination or infection, the body first produces antibodies of relatively low affinity for the foreign molecule, the antigen. As an immune response progresses, these antibodies become hypermutated. The purpose of hypermutation is to create new protein sequences that can bind the antigen more strongly and specifically than their precursors. This allows us to respond quickly

and effectively to pathogens that we have encountered previously. People who cannot mutate their antibodies suffer recurring bacterial and viral infections, and do not respond to vaccination.

Hypermutation occurs only in B cells that have been stimulated with antigen. During the development of B cells, before they encounter antigen, genetic shuffling has recombined blocks of DNA into long segments representing the basic antibody-encoding genes. Then, during hypermutation, many single mutations — and, rarely, small insertions and deletions — are introduced into the variable regions of these genes.

A breakthrough in understanding the mechanism behind hypermutation came two years ago, when Muramatsu *et al.*² discovered that the AID protein — which is expressed only in hypermutating B cells — is needed for the process. At the time it was thought that AID might convert ('deaminate') cytosine to uracil in the messenger RNA encoding an endonuclease enzyme, thereby enabling an active form of the enzyme to be produced. Endonucleases cut nicks in DNA strands, and it was proposed that nicks generated by this putative endonuclease in the variable regions of antibody genes could be the lesions that trigger hypermutation. This notion was attractive because numerous DNA breaks had been detected around variable genes. But the theory was questioned when similar numbers of breaks were found in mice lacking AID^{4,5}. Researchers now suspect that these nicks are unrelated to hypermutation.

A second possibility, proposed by Neuberger and colleagues, was that AID converts cytosine to uracil in DNA (not RNA), as suggested by the increase in mutations of

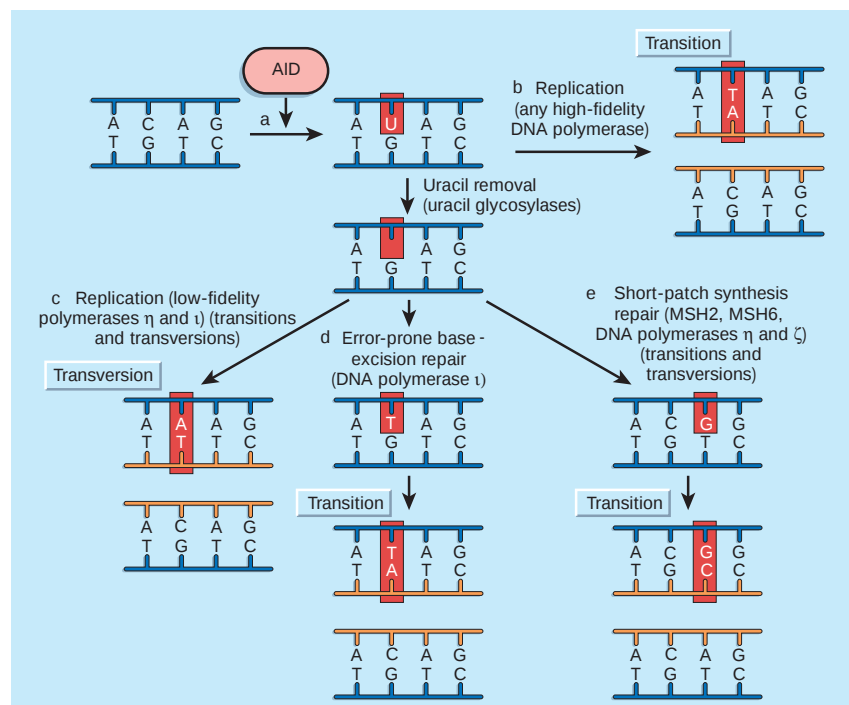


Figure 1 A model for antibody diversity. a, The AID protein converts cytosines (C) in the antibody-gene variable region to uracils (U). These can be processed in several ways to generate mutations. b, High-fidelity DNA-replicating enzymes (polymerases) recognize U as thymine (T), and pair it with adenine (A) during replication. c–e, These pathways rely on the initial removal of U by uracil glycosylases. c, Low-fidelity polymerases bypass the resulting gap during replication, filling it more or less at random. One example is shown. d, In this repair pathway, a T is preferentially inserted in the gap, leading to errors after replication. e, In this repair process, the gap is filled in and extended, generating mutations that are mostly located opposite As and Ts. In chicken B cells lacking the protein XRCC2, the replication pathways predominate. All four pathways work in mice and humans. ‘Transitions’ swap one purine base (G or A) for the other; likewise for pyrimidines (C, T and U). ‘Transversions’ swap a pyrimidine for a purine, or vice versa. Blue lines, old DNA strands; orange lines, new strands produced after DNA replication.

cytosines in AID-expressing bacteria³. In this model, AID produces a lot of uracils in the variable genes (Fig. 1a). The uracils are mistakes, as the usual bases in DNA are cytosine, guanine, adenine and thymine, and need to be corrected — but that often leads to more errors. One ‘correction’ mechanism involves the DNA polymerase enzymes that copy DNA, which mistakenly read uracil as thymine during replication, and introduce errors accordingly (Fig. 1b). Other correction mechanisms rely on the uracil first being removed by uracil glycosylase, leaving a gap. Mutations can then occur as the cell fills in the gap by one of three pathways — one DNA-replication and two DNA-repair pathways (Fig. 1c–e).

Di Noia and Neuberger¹ now provide support for the conversion of cytosine to uracil in DNA in eukaryotic cells. They took advantage of a chicken B-cell line engineered previously by Neuberger’s group. These cells lack XRCC2, a protein involved in DNA recombination and repair, and so tend to deal with uracil in DNA by the replication pathways (Fig. 1b, c), not by DNA repair. Moreover, they mainly generate transversion-type mutations in the variable genes. In transver-

sions, a pyrimidine base (cytosine, thymine or uracil) is swapped for a purine (adenine or guanine). In transitions, one pyrimidine is swapped for another, or one purine for another. The transversions in these B cells generally involve replacements of cytosine or guanine (rather than of adenine or thymine), and this probably occurs by the replication pathway that relies on the initial removal of uracil by the glycosylase (Fig. 1c).

The authors’ experiments were simple yet effective. The uracil glycosylase was inhibited in the B cells, forcing the cells to deal with uracils through the other replication pathway. By its nature this pathway generates transitions of an initial cytosine–guanine base pair to a thymine–adenine pair (Fig. 1b). Di Noia and Neuberger reasoned that if the production of uracil in variable-region genes is the trigger for hypermutation, and if cells cannot deal with the uracil by removing it but instead copy over it, then the predominant mutation pattern would shift from transversions of cytosine and guanine to transitions. And this is exactly what they saw.

So what about hypermutation in mammalian B cells? Again, the generation of uracil could be the crucial starting point. In

hypermutating mammalian B cells all four nucleotides are mutated with similar frequencies in variable genes, and transitions are favoured. All replication and repair pathways could be used to process AID-generated uracils, causing these patterns of hypermutation. After uracil is removed by either uracil glycosylase or another glycosylase, SMUG1, processing may be shunted into the repair pathways, allowing mutations of adenine and thymine — rather than just cytosine and guanine — to occur (Fig. 1d, e). During error-prone base-excision repair (Fig. 1d), the gap may be incorrectly filled in by DNA polymerase ι (ref. 6), which erroneously inserts thymine opposite guanine⁷. During short-patch repair synthesis (Fig. 1e), polymerase η could fill in the gap and synthesize several bases by strand displacement opposite adenines and thymine^{8,9}. The mismatch-repair proteins MSH2 and MSH6, and polymerase ζ , may also be involved^{10–12}.

The identification of AID-generated uracils as the likely starting point is a major breakthrough in understanding hypermutation. Nonetheless, questions remain. The most important one is how AID targets the rearranged variable genes. Perhaps AID is delivered to these genes by interacting with transcription factors¹³. Also, why is the usually accurate base-excision repair using polymerase β subverted to mutation-generating repair with polymerase ι in variable genes? What is the exact role of MSH2 and MSH6? AID also initiates two other antibody-alteration processes: ‘class switch recombination’ in mammals, and variable-gene conversion in chickens. How can the DNA-deaminase model explain these very different events? Finally, before the model is totally convincing it must be shown that AID actually does deaminate cytosine in DNA, and that uracils are present in the variable genes. But these are exciting times for immunology, and it appears that one of the last black boxes has been opened. ■

Patricia J. Gearhart is in the Laboratory of Molecular Gerontology, National Institute of Aging, National Institutes of Health, 5600 Nathan Shock Drive, Baltimore, Maryland 21224, USA.
e-mail: gearhartp@grc.nia.nih.gov

- Di Noia, J. & Neuberger, M. S. *Nature* **419**, 43–48 (2002); advance online publication, 31 July 2002 (doi:10.1038/nature00981).
- Muramatsu, M. et al. *Cell* **102**, 553–563 (2000).
- Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. *Nature* **418**, 99–103 (2002).
- Bross, L., Muramatsu, M., Kinoshita, K., Honjo, T. & Jacobs, H. *J. Exp. Med.* **195**, 1187–1192 (2002).
- Papavasiliou, F. N. & Schatz, D. G. *J. Exp. Med.* **195**, 1193–1198 (2002).
- Bebenek, K. et al. *Science* **291**, 2156–2159 (2001).
- Frank, E. G. et al. *EMBO J.* **20**, 2914–2922 (2001).
- Zeng, X. et al. *Nature Immunol.* **2**, 537–541 (2001).
- Pavlov, Y. I. et al. *Proc. Natl Acad. Sci. USA* **99**, 9954–9959 (2002).
- Phung, Q. H. et al. *J. Exp. Med.* **187**, 1745–1751 (1998).
- Wiesendanger, M., Kneitz, B., Edelmann, W. & Scharff, M. D. *J. Exp. Med.* **191**, 579–584 (2000).
- Diaz, M., Verkoczy, L. K., Flajnik, M. F. & Klinman, N. R. *J. Immunol.* **167**, 327–335 (2001).
- Peters, A. & Storb, U. *Immunology* **4**, 57–65 (1996).

Seismology

Stressed to quaking point

Chris Marone

The Earth's crust can deform catastrophically in earthquakes, but it's difficult to predict exactly what causes such failure. Analysing thousands of small shocks might help us better understand how earthquakes occur.

In the study of earthquake mechanics, one of the biggest problems is pinning down the initial conditions. The slip history and the initial stress field around a fault are rarely known with certainty, making it tricky at best to work out what conditions trigger an earthquake. Uncertainty over fault conditions limit hypothesis-testing and hamper both fundamental studies of earthquake physics and empirical attempts to predict earthquakes. Toda and colleagues¹ (page 58 of this issue) now report a new way to circumvent this problem. Rather than relying on estimates of the absolute stress state, they use the rate of stressing to study earthquake nucleation and seismic deformation. They find that regions stressed at higher rates experience more earthquakes in a given period, in agreement with fault-friction theory².

Toda *et al.* used a dense network of seismometers in the Izu Islands volcanic chain, south of Tokyo (Fig. 1), and a seismic catalogue extending back to 1980 to measure changes in the rate of earthquake occurrence around one of the most seismically energetic magma intrusions ever known³ — in 2000 the region was hit by a 'swarm' of more than 7,000 shocks. The authors infer that a vertical fracture, 8 km below the surface and 5 km × 15 km in area, was forcibly expanded to a width of 20 m by magma intrusion at an average rate of 300 m³ s⁻¹ during a two-month period. In some areas the local stressing rate increased by a factor of up to 1,500 relative to the background rate. By comparing seismic activity before and after the intrusion, the authors find that earthquake rates jumped by a factor of nearly 1,000 in the areas of highest stressing rate, some locations suffering daily the equivalent of 1,000 earthquakes of

magnitude 3 or more. They also show that the rate of seismic activity decreases in areas where the stressing rate decreases, in agreement with previous work⁴.

Significant in the work of Toda *et al.*¹ is their use of earthquake observations to test seismicity–rate theory² and the laboratory-derived friction laws⁵ on which it is based. The friction laws indicate that fault strength is not well described by a simple stress threshold but rather is a function of strain rate and recent slip history — that is, the 'frictional state'. History- or state-dependence of friction accounts for time-dependent strengthening of frictional contacts and for the fact that a finite displacement is necessary for the frictional resistance to make a transition from one set of conditions to another — for example, from stationary to sliding contact, or from steady sliding at one velocity or normal stress to another. Friction rate and state effects are well documented in laboratory experiments; however, testing how well they apply to earthquake faults has proved difficult and controversial.

At issue is whether the same processes that occur under laboratory conditions also govern seismic failure in the field, because the latter may involve slip rates of up to several metres per second and significant shear heating. Although the laboratory laws can reproduce much observed fault behaviour (including slow earthquakes, aseismic strain transients, dynamic rupture and interseismic fault healing^{5,6}), the stumbling block is that numerical simulations of these phenomena in the field must generally use parameter values that differ from those measured in the laboratory. Seismic and geodetic observations can rarely specify friction values with sufficient precision to resolve the discrepancy. But Toda *et al.* have devised a different type of test, combining field-based estimates of the friction parameters with laboratory-based predictions of seismic behaviour.

Dieterich's seismicity–rate theory² predicts that a stress perturbation vanishes sooner if the background stressing rate is higher. As aftershocks are a product of the stress perturbation generated by the main shock, their duration — the time for the seismicity to return to the pre-mainshock level — should scale inversely with stressing rate. The Izu Islands swarm includes five magnitude-6 earthquakes, and Toda *et al.* record a strong inverse correlation between aftershock

duration and stressing rate: aftershocks of magnitude-6 earthquakes, which would normally go on for several years, lasted only a day in the areas of highest stressing rate.

This agreement between theory and observation is a significant step forward for earthquake mechanics. The seismicity–rate theory also predicts a relationship between aftershock duration, stressing rate, and a friction parameter that describes the instantaneous friction response to a step change in slip velocity. Modelling studies show that larger values of the friction parameter prolong the time needed to reach instability⁷; hence, aftershock duration increases with the friction parameter. Laboratory experiments⁸ show that the parameter (which is always positive) is between 10⁻³ and 10⁻², in good agreement with the value obtained by Toda *et al.*¹ from measured aftershock durations and stressing rates.

Seismicity–rate theory also predicts that a change in stress level will have a markedly different effect from a change in stressing rate, and Toda *et al.* show how this could improve understanding of postseismic behaviour. The main shock of an earthquake increases the stress around a rupture, but this increase decays rapidly because the brittle crust behaves as a stiff elastic material. In contrast, the effective stiffness is expected to be significantly lower in some cases, including large earthquakes that rupture into the frictionally viscous region below the seismogenic zone. Toda *et al.* note that the combined effects of a jump in stress level and stressing rate could help explain transient behaviour observed after large earthquakes.

Given the issues discussed above, it is clearly difficult to predict details of a particular earthquake, including when and where it may nucleate. But the method of Toda *et al.* presents an opportunity to test earthquake theories that could lead the way to fundamental breakthroughs. The correlation between stressing rate and seismicity may help in forecasting swarm or aftershock damage; however, there is still the problem of rupture size and understanding how earthquakes stop. Because large earthquakes appear to be seismically identical to small shocks, forecasts and damage predictions will remain limited, at least for the moment, to minimum estimates.

Chris Marone is in the Department of Geosciences, 536 Deike Building, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

e-mail: cjm@essc.psu.edu

1. Toda, S., Stein, R. S. & Sagiya, T. *Nature* **419**, 58–61 (2002).
2. Dieterich, J. H. *J. Geophys. Res.* **99**, 2601–2618 (1994).
3. <http://sdc.wr.usgs.gov>
4. Dieterich, J. H., Cayol, V. & Okubo, P. *Nature* **408**, 457–460 (2000).
5. Marone, C. *Annu. Rev. Earth Planet. Sci.* **26**, 643–696 (1998).
6. Scholz, C. H. *Nature* **391**, 37–42 (1998).
7. Roy, M. & Marone, C. *J. Geophys. Res.* **101**, 13919–13932 (1996).
8. Blanpied, M. L., Marone, C., Lockner, D. A., Byerlee, J. D. & King, D. P. *J. Geophys. Res.* **103**, 9691–9712 (1998).

Figure 1 Miyakejima, Izu Islands, August 2000. Mount Oyama erupts as the islands suffer thousands of earthquakes.

A lost Neanderthal neonate found

A remarkable discovery in a French museum answers some long-standing questions.

Fossil remains of adult Neanderthals are well documented, but juvenile specimens are rare and information about them is scant. Here we identify a beautifully preserved skeleton that has been lost to science for almost 90 years as the Neanderthal neonate known as 'Le Moustier 2', which was originally found at Le Moustier in the Dordogne, southwest France. This find will be a

rich source of data for studying the evolution of human ontogeny¹ as well as the phylogenetic relationship between these extinct hominids and anatomically modern humans.

The burial site of Le Moustier 2 was discovered at the lower rock shelter in Le Moustier by Peyrony in 1914 (ref. 2). The burial pit originated in archaeological level J (ref. 3), a Mousterian cultural layer that has been dated to $40,300 \pm 2,600$ years BP by thermoluminescence⁴. Although at the time the bones were thought to belong to a neonate^{2,3}, the specimen was not investigated further and the remains were thought to have been lost in Paris⁵.

During a survey in 1996 of collections in the Musée National de Préhistoire in Les Eyzies-de-Tayac-Sireuil, France, the remains of a neonatal skeleton were found among the lithic sample from Le Moustier. Some bones were isolated, but others were still embedded in blocks of sediment. The possibility that these might be the bones of Le Moustier 2 prompted a re-examination of Peyrony's notes, which made no mention of the specimen ever having been sent to Paris.

Several crucial details emerged during the cleaning and restoration of the specimen (Fig. 1) that confirm that this skeleton is indeed the missing Le Moustier 2. The fossil was embedded in a sediment of sand containing green hornblendes, garnets and mica particles of muscovite type, a composition that has always been associated with the deposition of the Vézère river, which flows past the Le Moustier cliff. Preliminary sedimentological analysis revealed many similarities with layers I and J of Le Moustier. Furthermore, flint flakes and faunal remains (*Rangifer tarandus*, *Cervus* sp., *Capra hircus ibex* and a large Bovinae) in the sediment around the skeleton are identical to those found in the Mousterian levels at the Le Moustier inferior rock shelter. Estimations of stature based on the length of the long bones⁶ confirmed that the age at death was no more than four months.

Although Le Moustier 2 is one of the most complete Neanderthal individuals to have been discovered, both scapulae and the pubis are missing. However, the right femur and right humerus, which were previously also missing, have now been found, having been wrongly attributed to the Neanderthal La Ferrassie 4 (LF4).

It had been assumed that the right femur and right humerus came from the burial pit of the La Ferrassie 4bis Neanderthal (LF4bis), even though their colour and fossilization differ markedly from those of LF4bis (ref. 7). Moreover, the sediment

adhering to the LF4 bones contains muscovite, a mineral that is not present at La Ferrassie but which is prevalent at Le Moustier. Besides the matching fossilization and colour, the bones' morphology, variability, muscular insertions, discrete traits and measurements are identical to those of the left humerus and femur of the specimen from the Musée National de Préhistoire.

It is likely that the two bones attributed to LF4 are in fact samples from Le Moustier 2 that were sent by Peyrony to Boule for an age diagnosis. The attribution of the LF4 limb bones to Le Moustier 2 eliminates from the archaeological record the only double burial associated with Neanderthals.

The morphology of Le Moustier 2 differs greatly from that of living human neonates, but shows many similarities with the features of juvenile and adult Neanderthals. For example, the left stapes has asymmetrical limbs⁷; there is no infra-orbital depression on the maxilla⁸; and the premaxillary suture is fully open on both the palatal surface (accompanied by two interincisive sinuses) and the nasal floor⁹. Also, the zygomatic bone is short in comparison with the length of the frontal process, as in immature Neanderthals¹⁰; the sagittal profile of the nasal bones is the same as that in adult Neanderthals^{9,11}; and in the dentition, the crowns of the central and lateral upper deciduous incisors are shovel-shaped and biconvex, features that are often evident in Neanderthals¹².

This taxonomic diagnosis awaits final confirmation through atomic mass spectroscopic carbon-dating of the skeleton. Nevertheless, the differences between the modern human skeleton and that of Le Moustier 2 bear out the proposed extensive genetic variation between Neanderthals and extant humans¹³.

Bruno Maureille

UMR 5809 CNRS, Laboratoire d'Anthropologie des Populations du Passé, Université Bordeaux,
1 avenue des Facultés, 33405 Talence, France
e-mail: b.maureille@anthropologie.u-bordeaux.fr



Figure 1 The neonatal skeleton Le Moustier 2, one of the most complete Neanderthal individuals ever discovered. The parietals and the posterior part of the hemifrontals are broken into small fragments. The left temporal, both scapulae and the pubis are missing. Most of the cranial base and face, deciduous tooth germs, cervical vertebrae and long bones are preserved. The right humerus and femur, formerly misidentified as La Ferrassie 4, are not shown. Many Neanderthal derived traits can be seen in the lateral and basal portions of the occipital bones and the petrous portion of the temporal bone. Derived traits are also evident in the postcranial skeleton in the medio-lateral curvature of the radius, the relative proportion of the thumb phalanges, the orientation of the ulna head and the curvature of the ribs. The thickness and mass of the bones are noteworthy.

1. Ponce de Léon, M. S. & Zollikofer, C. P. E. *Nature* **412**, 534–538 (2001).
2. Peyrony, D. *Rev. d'Anthropol.* **40**, 48–76, 155–176 (1930).
3. Peyrony, D. *Les Moustériens Inhumains: Ils Leurs Morts?* (Ribes, Périgueux, 1921).
4. Valladas, H. et al. *Nature* **322**, 452–454 (1986).
5. Heim, J.-L. *Les Hommes Fossiles de La Ferrassie* Vol. 1 (Masson, Paris, 1976).
6. Sellier, P., Tillier, A.-M. & Bruzek, J. *Am. J. Phys. Anthropol.* **22** (suppl.), 208 (1997).
7. Heim, J.-L. *Les Enfants Néandertaliens de La Ferrassie* (Masson, Paris, 1982).
8. Maureille, B. *La face chez Homo erectus et Homo sapiens: Recherche sur la variabilité morphologique et métrique*. Thesis, Univ. Bordeaux 1, France (1994).
9. Maureille, B. & Bar, D. *J. Hum. Evol.* **37**, 137–152 (1999).

10. Dodo, Y., Kondo, O., Muhsen, S. & Akasawa, T. in *Neanderthals and Modern Humans in Western Asia* (eds Akasawa, T., Aoki, K. & Bar-Yosef, O.) 323–338 (Plenum, New York, 1998).
11. Trinkaus, E. *The Shanidar Neanderthals* (Academic, New York, 1983).
12. Patte E. *La Dentition des Néandertaliens* (Masson, Paris, 1962).
13. Krings, M. et al. *Nature Genet.* **26**, 144–146 (2000).

Competing financial interests: declared none.

Ecology

Effect of British hunting ban on fox numbers

Pressure to ban the hunting of foxes with hounds in Britain has fuelled debate about its contribution to the control of fox populations. We took advantage of a nationwide one-year ban on fox-hunting during the outbreak of foot-and-mouth disease (FMD) in 2001 to examine this issue and found that the ban had no measurable impact on fox numbers in randomly selected areas. Our results argue against suggestions that fox populations would increase markedly in the event of a permanent ban on hunting.

In the 2000–2001 season, foxes were hunted (Fig. 1) by 196 registered foxhound packs and 7 harrier packs¹, and by an unknown number of unregistered packs, the latter mainly in southwest England and Wales². However, their contribution to controlling fox numbers is unclear. The arguments are that the number of foxes killed by hunting is relatively small, and therefore that hunting has little influence in regulating population size³, and conversely that hunting is important for limiting fox numbers, particularly in

combination with other culling practices, so a ban would lead to a significant population increase with consequential economic damage and loss of biodiversity⁴.

The main fox-hunting season in Britain runs from November to March/April; cub-hunting lasts from August to October, and ‘problem’ foxes are targeted after the main season. From 23 February to 17 December 2001, the Foot-and-Mouth Disease Declaration (Controlled Area) Order 2001 banned hunting. From 17 December, some hunts were licensed to hunt in FMD-free areas, and from 11 February 2002, any hunt could be licensed if certain restrictions were observed. Hunting was thus banned for ten months and severely curtailed for two further months.

With 240,000 adult foxes, and 425,000 cubs born in Britain each year⁵, 64% mortality per annum is required to prevent growth of the fox population⁶. A one-year ban should therefore lead to a detectable population increase if hunting has a significant impact on fox numbers.

To quantify the changes in fox numbers during the hunting ban, we surveyed 160 one-kilometre squares between 1 February and 17 March in 1999 and 2000 (pre-FMD) and 2002 (post-FMD) — that is, towards the end of the main culling period and just before the cubs are born. We estimated relative fox density by using faecal-accumulation rates⁷ along transects (mean length, 6.9 km) that followed linear features; we surveyed each transect twice in each period (pre- and post-FMD). We removed all fox faeces on the first visit, and counted fresh faeces 2–6 weeks later. We calculated faecal density for each transect in each period (F) as $F = S/KD$, where S is the number of faeces on the second visit, K is the transect length in kilometres, and D is the number of days between visits. Absolute changes in faecal density were calculated as $F_{\text{post-FMD}} - F_{\text{pre-FMD}}$, and relative changes (R') as $\log(R + 1)$, where $R = (F_{\text{post-FMD}} - F_{\text{pre-FMD}}) / (F_{\text{pre-FMD}} + 1)$.

Overall, faecal density declined by 4.7% (Fig. 2a), with no significant difference in the number of squares where faecal density increased ($n = 83$) compared with those that decreased or remained unchanged ($n = 77$) in ‘hunted’ ($n = 118$) as opposed to ‘not-hunted’ ($n = 42$) squares ($\chi^2 = 0.19$, $P = 0.66$). Moreover, neither the absolute ($t_{158} = 0.60$, $P = 0.55$) nor the relative ($t_{158} = 0.51$, $P = 0.61$) change in faecal density differed between ‘hunted’ and ‘not-hunted’ squares. ‘Hunted’ squares lay within the area covered by a hunt, as designated on a map issued in 1996 (ref. 8).

R' values were normally distributed



Figure 1 Hunting with hounds. This centuries-old British tradition may not be particularly effective in controlling fox populations.

(Kolmogorov–Smirnov goodness-of-fit test, $Z = 0.71$, $P = 0.70$), with homogenous variances between regions (Levene statistic, $F_{8,151} = 1.84$, $P = 0.07$). There was a significant difference between nine regions in the relative magnitude of the change in faecal density (ANOVA, $F_{8,145} = 2.69$, $P = 0.01$). Tukey’s HSD *post hoc* comparisons showed a significant increase in eastern England (one sample t -test, $t_{22} = 2.52$, $P = 0.02$) and a significant decrease in southern England ($t_{37} = -2.65$, $P = 0.01$). No other region deviated significantly from zero in test values.

We evaluated the relative change in faecal density against a regional index of hunting pressure (number of days hunted by registered packs per week per km² in the 2000–2001 season¹). There was no association between the reduction in hunting pressure in each region and the relative change in faecal density (linear regression with replication, $r^2 = 0.007$, $F_{1,158} = 1.18$, $P = 0.28$; Fig. 2b). We conclude that there was no significant change in fox numbers during the one-year hunting ban, and that in most regions the average faecal density had declined. Our results therefore support the view taken by the Committee of Inquiry into Hunting with Dogs² that a permanent ban on hunting is unlikely to result in a dramatic increase in fox numbers.

Philip J. Baker, Stephen Harris, Charlotte C. Webbon

School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK

e-mail: s.harris@bristol.ac.uk

1. Anon. *Baily’s Hunting Directory 2000–2001* (Pearson, Cambridge, 2000).
2. Burns, L., Edwards, V., Marsh, J., Soulsby, L. & Winter, M. *Report of the Committee of Inquiry into Hunting with Dogs in England and Wales* (Stationery Office, London, 2000).
3. Macdonald, D. W. & Johnson, P. J. in *The Exploitation of Mammal Populations* (eds Taylor, V. J. & Dunstone, N.) 160–207 (Chapman & Hall, London, 1996).
4. Tapper, S. *A Question of Balance: Game Animals and their Role in the British Countryside* (Game Conservancy Trust, Fordingbridge, Hampshire, 1999).
5. Harris, S., Morris, P., Wray, S. & Yalden, D. *A Review of British Mammals: Population Estimates and Conservation Status of British Mammals other than Cetaceans* (Joint Nat. Conserv. Commit., Peterborough, 1995).
6. Harris, S. & Saunders, G. *Symp. Zool. Soc. Lond.* **65**, 441–464 (1993).
7. Putman, R. J. *Mammal Rev.* **14**, 79–97 (1984).
8. Alexander, K. *Baily’s Hunting Directory 1996–1997* (Pearson, Cambridge, 1996).

Competing interests: financial, declared none; P.J.B. and C.C.W. were funded by the International Fund for Animal Welfare.

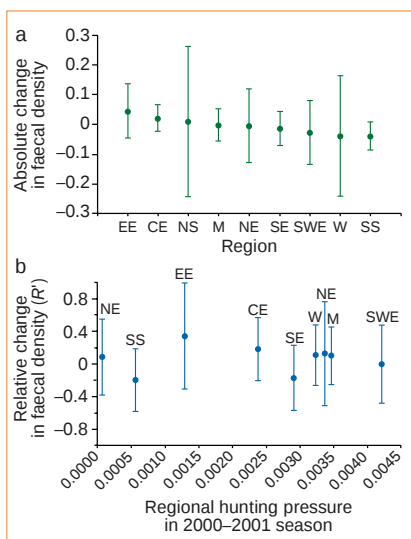


Figure 2 Changes in faecal density before and after the ban on fox-hunting during the 2001 outbreak of foot-and-mouth disease (FMD) in Britain. **a**, Mean ($\pm$ s.d.) change in absolute faecal density by region (post-FMD minus pre-FMD). **b**, Relationship between regional reduction in hunting pressure and relative change in faecal density. CE, central England, $n = 16$; EE, eastern England, $n = 23$; M, Midlands, $n = 21$; NE, northern England, $n = 15$; NS, northern Scotland, $n = 8$; SE, southern England, $n = 38$; SS, southern Scotland, $n = 9$; SWE, southwest England, $n = 25$; W, Wales, $n = 5$. The results of this survey are posted by The Mammal Society at www.mammal.org.uk.

The voltage-gated potassium channels and their relatives

Gary Yellen

Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA

The voltage-gated potassium channels are the prototypical members of a family of membrane signalling proteins. These protein-based machines have pores that pass millions of ions per second across the membrane with astonishing selectivity, and their gates snap open and shut in milliseconds as they sense changes in voltage or ligand concentration. The architectural modules and functional components of these sophisticated signalling molecules are becoming clear, but some important links remain to be elucidated.

Electricity plays an unavoidable role in biology. Whenever solutes like phosphate compounds, amino acids, or inorganic ions are transported across membranes, the movement of their charge constitutes an electrical current that produces a voltage difference across the membrane. Beyond managing to keep this accumulation of charge from getting too far out of balance, all living cells have developed the ability to exploit a transmembrane electrical potential as an intermediate in the storage of energy and the synthesis of ATP. However, a specialized class of animal cells—neurons, muscle cells and endocrine cells; known collectively as the ‘excitable cells’—have made the management and production of electrical signals into a high art.

Fast electrical signalling is made possible by the slow and steady homeostatic mechanisms that establish the standard environment and content of animal cells: high Na^+ concentration ($[\text{Na}^+]$) in the blood and extracellular fluid, and high $[\text{K}^+]$ (but low $[\text{Na}^+]$ and $[\text{Ca}^{2+}]$) in the cytoplasm. The plasma membrane separates these two compartments. These gradients are established and maintained by active transporters and pumps, which prepares the way for rapid changes in membrane voltage to be produced by passive transport through ion channels. These pore-forming proteins allow ions to flow only ‘downhill’, as dictated by their electrochemical gradients, but they do it rapidly and selectively. Opening a Na^+ -selective channel permits Na^+ to flow down its gradient into a cell, making the intracellular voltage more positive. Opening a K^+ -selective channel permits K^+ to flow out of a cell and restores the voltage to a negative value. The dimensions of a typical cell and its membrane allow the voltage to be changed rapidly back and forth many times, with relatively small changes in concentration. This is essentially how all cellular electrical signalling is produced.

The ultimate specialization of electrical signalling is the action potential, a stereotyped, millisecond-long electrical signal that is capable of propagating at rates of metres per second along a nerve fibre. As Hodgkin and Huxley showed¹, this signal is made possible by a rapid feedback process involving a direct action of voltage on ion channels: Na^+ channels activated (opened) by positive voltage (constituting the positive feedback process), followed by K^+ channels activated more slowly by positive voltage (the negative feedback).

Basic channel architecture

The voltage-gated K^+ channels are the prototypical voltage-gated channels. At their simplest, they are homotetrameric channels, with each subunit containing a voltage sensor and contributing to the central pore. The standard K^+ channel subunit contains six transmembrane regions, with both amino and carboxy termini on the intracellular side of the membrane (a tetrameric 6TM architecture; see Box 1). The pore-forming subunits of voltage-gated Na^+ and

Ca^{2+} channels contain four non-identical repeats of this motif, strung together on a single polypeptide. There is enormous variety within each of these channel families: voltage-gated K^+ channels alone are made by at least 22 different genes in mammals, with additional variety produced by alternative splicing and heteromultimerization.

Some close relatives of the voltage-gated K^+ channels share the tetrameric 6TM architecture but differ in key functional features. These include the sensory channels of the photoreceptors and olfactory neurons, which are non-selective (Na^+ and K^+) channels that are cyclic-nucleotide-gated (the CNG channels) and insensitive to voltage, and the pacemaker channels found in heart muscle and neurons (the HCN channels) that are controlled by both cyclic nucleotides and voltage.

Each of these signalling proteins performs three main functions. Ion permeation is the central function, and the movement of ions through the pore must be fast and ion-selective. This permeation is then regulated by opening and closing of the pore—a set of conformational changes called gating. Finally, the gating is coupled to a sensing mechanism, which detects transmembrane voltage in the core members of the family, but can also be geared to sense Ca^{2+} , cyclic nucleotides, and perhaps other cellular signals. From approximately 40 years of electrophysiology and biophysics, a little over a decade of cloning followed by mutagenesis and physiology on cloned channels, and from a few important crystal structures of some bacterial relatives of these channels, we now have a relatively clear picture of how these functions are accomplished.

Exquisite selectivity with high throughput

K^+ channels are extremely selective in which ions they allow to pass, yet they allow transport rates close to the aqueous diffusion limit. Selectivity and speed are both crucial for the biological function of the channels. As discussed above, in neurons and other cells, the ion specificity of a channel (for example, Na^+ or K^+) determines the direction of current flow. Na^+ and Ca^{2+} ions flow inward and thus carry positive charge into the cell—because most voltage-activated channels are activated by positive voltage, this further activates channels and produces regenerative excitation. K^+ ions flowing outward (or Cl^- ions flowing inward) reduce the positive charge inside the cell, and terminate (or prevent) excitation.

High flow rates are essential for producing these voltage changes rapidly. A typical action potential in a mammalian neuron requires millions of ions to flow in a millisecond. To accomplish this without using millions of proteins requires that each one have a very high throughput, while maintaining high selectivity.

What are the energetic and structural requirements for rapid permeation? The reason that ions require assistance to cross lipid

Box 1

The tetrameric 6TM architecture of the K⁺ channel family

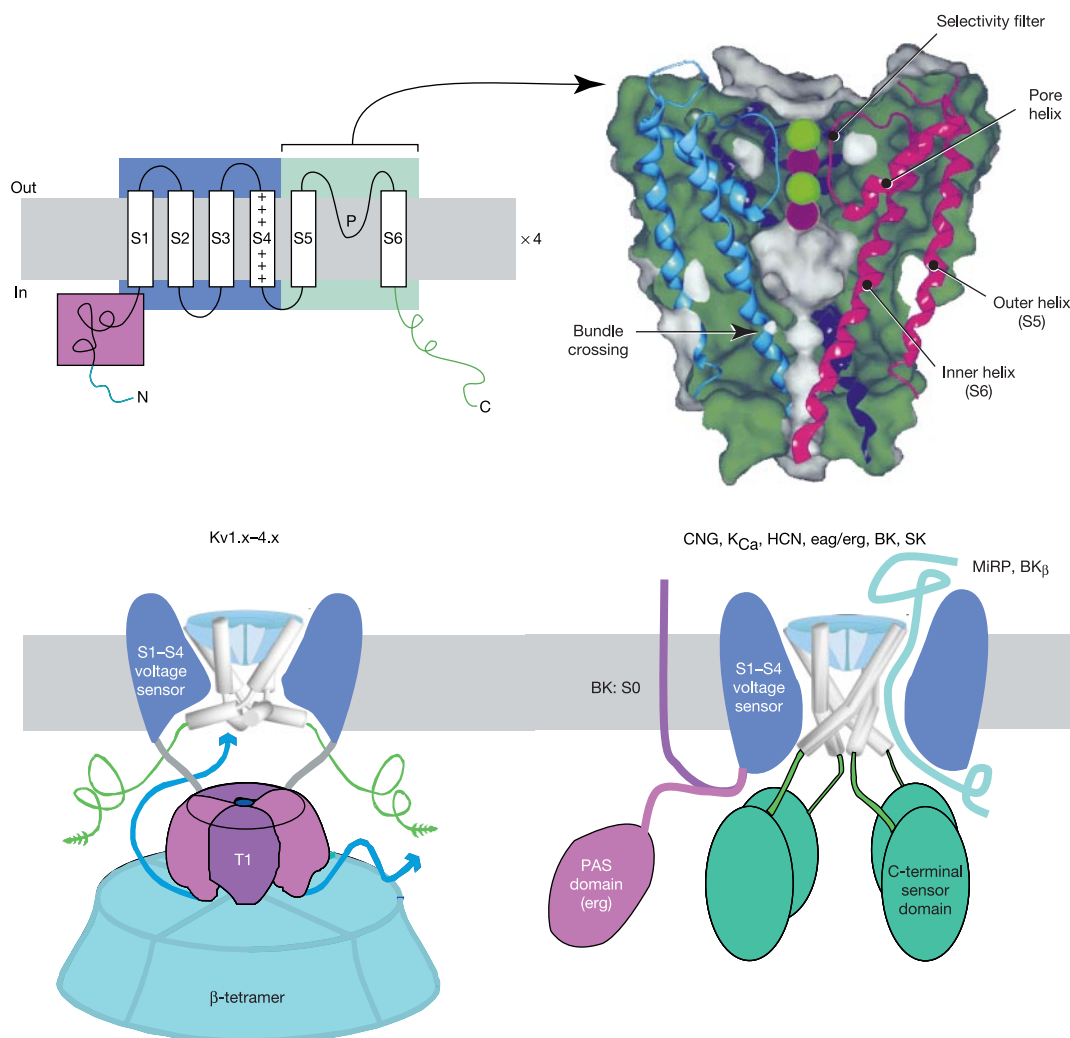
The typical voltage-gated K⁺ channel is an assembly of four identical (or similar) transmembrane subunits surrounding a central pore. Each subunit has six transmembrane crossings (S1–S6), with both N- and C termini on the intracellular side of the membrane (top left panel). The narrowest part of the pore, the selectivity filter, is formed by a loop between S5 and S6; the voltage sensor includes the S4 region with its multiple positive charges.

The bacterial KcsA K⁺ channel (top right panel, in cross-section) is the prototype for the pore-forming domain of the channel. It has only two transmembrane helices: the outer helix is homologous to S5 and the inner helix to S6. The interior of the protein is dark green, with secondary structure shown as ribbons for three of the four subunits, and the water-exposed surface of the protein is grey. The narrow selectivity filter is seen at the top, with the extended selectivity filter loop supported by the pore helices. The four spheres mark the four K⁺ ion binding sites; these are typically occupied by alternating K⁺ ions and water molecules⁵. The four inner helices cross to constrict the channel at the bottom (the 'bundle crossing'). Between the selectivity filter and bundle crossing is the water-filled 'cavity'.

In the K⁺ channel family there are two basic patterns for organizing the structural domains of the channels. All have the central 6TM structure, with a pore domain formed by S5–P–S6 and a voltage sensor comprising S1–S4. The core Kv1.x–Kv4.x subfamily (lower left panel) has at its N terminus a 'tetramerization domain' (T1, purple)^{3,74} that determines the specificity of subunit assembly⁷⁵ and also serves as a

platform for attachment of the optional β -subunits⁷⁴ and for other protein–protein interactions. These interactions probably modulate channel activity and may also provide an avenue for the channels to signal directly to intracellular signalling processes⁷⁶. The extreme N terminus can provide the auto-inhibitory 'ball' for channel inactivation (see text). The C termini (green) have no obvious domain structure in this subfamily, although there is commonly a PDZ-binding motif⁷⁷ that determines the physical localization of the channel and its association into large signalling complexes.

Many other tetrameric 6TM channels have a different domain structure, characterized by the absence of a T1 domain and the presence of a sensor domain in the C terminus (lower right panel). These include the voltage-gated KCNQ and eag/erg channels, as well as channels gated coordinately by voltage and ligand binding (BK Ca²⁺-activated channels and HCN hyperpolarization and nucleotide-gated channels), and channels gated exclusively by intracellular ligands (CNG channels and SK channels). In some cases the C-terminal sensor domains adopt a four-fold symmetrical organization⁴⁷, whereas in other cases they may function as a dimer of dimers⁵². At the N terminus some family members have an additional transmembrane region (S0)⁷⁸ or an additional sensor domain⁷⁹. The auxiliary subunits known to be associated with this family (minK, MiRP, BK β) are transmembrane; they appear to be intimately associated with the pore domain^{80,81} and may displace the voltage-sensor domain or interact with it directly.



membranes is that they have an extremely favourable interaction with water. In the vicinity of a positively charged ion like K^+ , the water molecules re-orient with their negatively charged end towards the ion. In lipid or even in protein, there is no analogous electrostatic accommodation to the presence of the ion (these substances are said to have a relatively low dielectric constant compared with water).

One option for making the K^+ ion stable inside a protein pore would be to use a negative charge. The risk in doing so is that the K^+ ion might become too attracted to the negative charge, and wouldn't exit the pore, as required for a high throughput. This is not the mechanism used by K^+ channels.

Instead, the K^+ channel proteins have four specific architectural features that keep the ion almost exactly as stable as it is in water. First, the channels use plenty of water to make the ion stable. Rather than maintaining a narrow pore of atomic dimension through the entire thickness of the membrane, part of the permeation pathway is broad and contains a large amount of water (see Fig. 1). This explanation was proposed many years ago, particularly as an explanation of the very large conductance of certain Ca^{2+} -activated K^+ channels², and is borne out in the remarkable crystal structures of the pores of two different bacterial K^+ channels, KcsA and MthK^{3–6}.

Second, the K^+ channels apparently stabilize the ions and achieve cation selectivity by using the electrostatic influence of helix dipoles. Every α -helix has a dipole moment owing to the alignment of the dipoles of its hydrogen bonds, and the intracellular vestibule of K^+ channels has the negative ends of four helix dipoles pointed towards its heart³. It has been proposed^{3,7} that these helix dipoles produce a preferential stabilization of cations near the entrance to the narrow selectivity filter. This concept is mirrored in the quite different architecture of channels designed to conduct the negatively charged chloride ions—these have the positive ends of multiple helices pointed towards the central ion site⁸.

A third feature of the K^+ channel's approach to permeation and ion selectivity is creation of a series of customized polar oxygen cages. As a K^+ ion diffuses through water itself, it is more or less constantly surrounded by a cage of polar oxygen atoms from the

water molecules. Water is extremely flexible and accommodating, and easily adapts the size and configuration of the cage to accommodate ions of different sizes, such as Na^+ and K^+ . The selectivity filter of the potassium channel is designed to mimic the water structure around a K^+ ion but is also designed not to adopt the presumably more compact structure around a Na^+ ion (as originally proposed by Armstrong⁹). Each K^+ ion in the selectivity filter is surrounded by two groups of four oxygen atoms, just as in water: these oxygen atoms are held in place by the protein, and are in fact the backbone carbonyl oxygens of the selectivity filter loops from the four subunits.

Finally, a long-known feature of potassium channels is that K^+ ions pass through in single file, with simultaneous occupancy by multiple ions^{5,10–12}. The mutual electrostatic repulsion between adjacent K^+ ions (spaced about 7 Å apart) destabilizes ions in the pore, permitting the other favourable interactions to produce ion selectivity without producing overly tight binding that would impair rapid permeation.

The structural and architectural features of potassium channels are thus perfectly adapted to fit their function. They solve the electrostatic problem of stabilizing ions—without making them too much more stable than they are in water—by using plenty of water and helix dipoles to counteract the unfavourable dielectric environment within the membrane. Furthermore, they solve the problem of stabilizing potassium in preference to sodium by precisely matching the configuration of oxygen atoms around a solvated potassium ion.

Gating the pore with conformational motions

To use these transmembrane pores for electrical signalling or even for regulation of ion transport, it is necessary to regulate their permeability. On the slowest timescales, it is possible to control the manufacture of channel proteins by a transcriptional regulatory mechanism; on a more rapid timescale, it is possible to insert multiple channels into the membrane (or withdraw them) by use of a vesicle fusion mechanism. But the very rapid signalling in neurons requires a fast mechanism for closing and opening the pore. There are three established mechanisms by which the voltage-gated K^+ channels can close: two of them involve a conformational constriction of the permeation pathway, and one involves conditional plugging of the pore by an auto-inhibitory part of the channel protein.

Gating at the S6 bundle crossing

First, the channel can close by pinching shut at the intracellular entrance. This intracellular or S6 gate obstructs entrance from the cytoplasmic surface to the water-filled 'cavity' in the centre of the channel protein. The S6 transmembrane region corresponds to the 'inner helix' of the bacterial KcsA channel. The top (extracellular) ends of the four S6 helices form a cradle for the selectivity filter, whereas the bottom ends converge to a right-handed 'bundle-crossing' below the cavity.

This bundle-crossing corresponds to the functionally determined closure point of at least some of the voltage-gated K^+ channels^{13,14}. The S6 transmembrane region and its intracellular extension show very high sequence conservation within the principal families of voltage-gated K^+ channels, and at the bundle crossing there is a strongly conserved proline sequence (PxP or PxG) that is not found in the bacterial K^+ channels. Studies of this region indicate that it is likely to bend the S6 in both the closed and open channels^{14–16}.

The probable nature of the S6 opening motion has now been made apparent by the determination of the structure of a second bacterial K^+ channel, MthK⁶, a tetrameric 2TM that can be opened by intracellular Ca^{2+} . This structure appears to represent the MthK channel in the open state (Fig. 2, centre panel), and it has no bundle crossing: instead, the homologues of the S6 helices are splayed wide open with no constriction between the intracellular solution and the

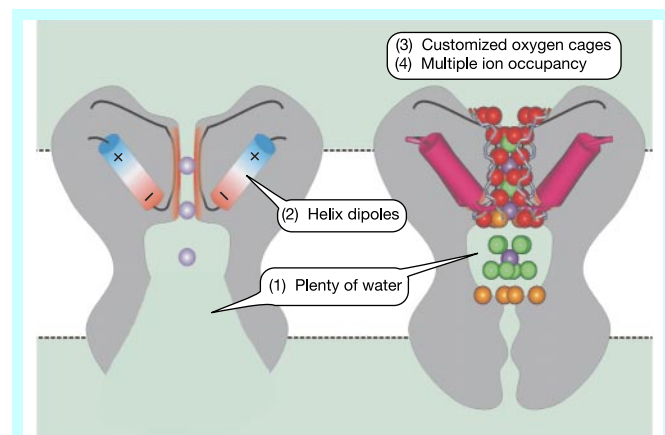


Figure 1 Architectural features of K^+ channels important for ion permeation.

Approximate cross-section of an open K^+ channel, based on the crystal structure of the MthK channel⁶ (left). The wide-open intracellular vestibule and pore helix dipoles are highlighted. The diagram on the right is a view derived from the high resolution structure of the KcsA channel⁴, showing a narrower and probably closed access to a water-filled cavity in the middle of the membrane protein. A trapped K^+ ion (purple sphere) and its eight water molecules of hydration (green spheres) are remarkably well resolved; the backbone oxygens of the selectivity filter (red spheres) provide a good match to this hydration environment for the K^+ ions. Occupancy of the pore by multiple K^+ ions has an important role in selectivity and other properties of the channel^{5,10–12}. This basic architecture can also produce non-selective cation channels²¹ or channels selective for Na^+ or Ca^{2+} ions²². Orange spheres indicate side chain oxygen atoms.

selectivity filter⁶. Comparing this with the presumably closed structure of KcsA gives a plausible picture for the gating motion (Fig. 2, top and centre panels): the S6 helices swing open from an apparent hinge located at a highly conserved glycine residue. The implied motion is much larger than that previously suggested from electron paramagnetic resonance (EPR) measurements on KcsA¹⁷. It is still unknown whether the voltage-gated channels may use the PxP sequence as a second (or alternative) hinge, or whether the PxP sequence provides a fixed bend that allows the S6 gate of these channels to connect with the transmembrane voltage sensor (rather than to the intracellular calcium sensor used by MthK).

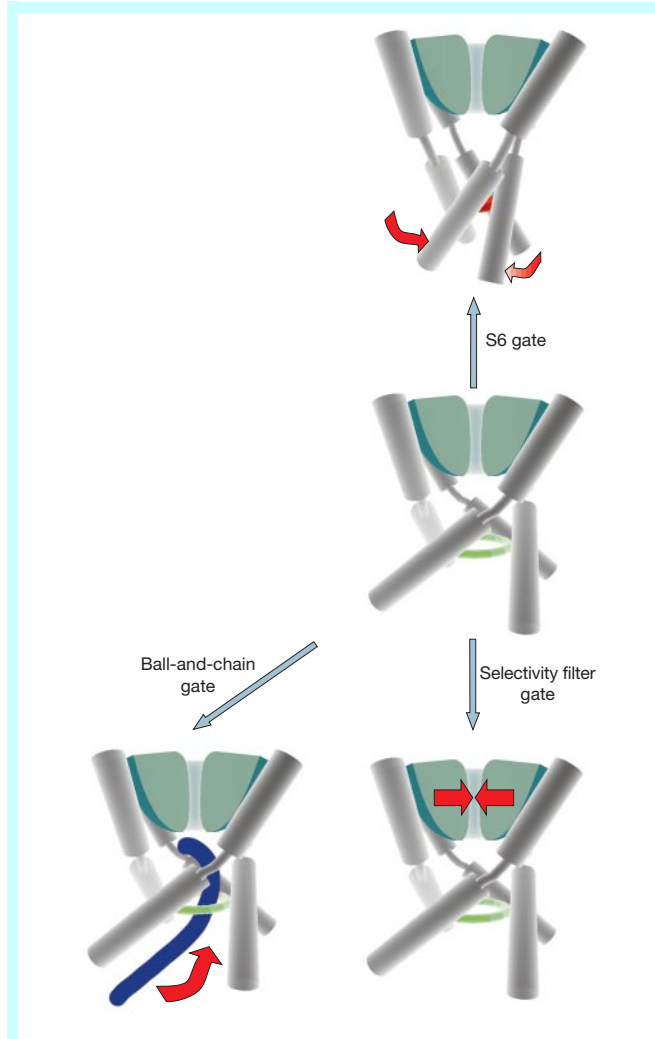


Figure 2 The conformational changes that gate the K^+ channel pore. The three best-understood conformational changes for closing potassium channels are shown here. The centre diagram shows the core structure of an open K^+ , on the basis of the MthK structure⁶. The selectivity filter (green) is shown with a cutaway to reveal the narrow pore. The outer helices have been stripped away to show the configuration of the inner helices, corresponding to S6. In the open state these are splayed apart, as in the MthK structure, using the conserved glycine hinge near the selectivity filter. The top diagram shows a closed K^+ , on the basis of the KcsA structure³. The four S6 transmembrane segments swing together to produce a secure closure at the bundle crossing¹⁴. This shuts off ion flow and can trap organic channel blockers in the enclosed cavity. The green and red rings in the diagrams illustrate the open and closed states, respectively, of the S6 gates. Alternatively, the channel may be closed at the selectivity filter (lower right diagram), or by binding of an auto-inhibitory peptide (lower left diagram) when a binding site in the pore is revealed by opening of the S6 gate. This figure was prepared with WebLab Viewer (<http://www.accelrys.com>), AC3D (<http://www.ac3d.org>), and POV-Ray (<http://www.povray.org>).

An important consequence of the S6 gating mechanism is its interaction with channel blockers. Small organic blockers, including local anaesthetic-like molecules, act by binding to sites within the cavity. Because of the S6 gate, these blockers can enter the channel only after the channel has been opened (by voltage or another stimulus), and some blockers can be trapped in the cavity when the S6 gate closes^{18,19}. The early discovery of these phenomena by Armstrong inspired the initial hypothesis of an intracellular gate^{9,18}. Blocker gating and trapping is also a primary mechanism for state-dependent and use-dependent action of channel blockers, a feature that is often crucial to the therapeutic value of channel drugs (for instance, anticonvulsants that must inhibit excessive electrical activity without affecting normal lower levels of activity).

Ball-and-chain gating

A second mechanism for closing the pore uses the S6 gate to regulate the binding of an auto-inhibitory peptide that is part of the channel protein (Fig. 2, lower right panel). Similar to the small organic blockers studied by Armstrong^{9,18}, the N terminus of Shaker K^+ channels can act as a channel blocker, probably by binding directly to the cavity. Because this interaction occurs only after the S6 gate has opened, some voltage-gated K^+ channels conduct only transiently, during the brief period after the S6 gate opens and before the N-terminal peptide blocks the channels: electrophysiologists call this 'N-type inactivation'.

N-type inactivation (also called the ball-and-chain mechanism because it involves a tethered blocker, initially imagined to be a ball) occurs in several known voltage-gated K^+ channels, and can involve either the N terminus of the principal (α) subunit of the channel²⁰ or the N terminus of an associated β -subunit²¹. It can be disrupted by enzymatic or genetic removal of the N-terminal sequence²⁰, and restored by addition of a soluble peptide of the same sequence to the intracellular bathing solution²². There is little sequence conservation between the various N termini capable of producing this form of inactivation, although both positive charge and hydrophobic character are important for the interaction with the open channel^{23,24}. The peptide appears to inhibit the channel by a simple blocking mechanism: it competes with intracellularly applied channel blockers²⁵, is expelled by potassium flow from the opposite side of the membrane²⁶, and is specifically affected by mutation of residues within the cavity²⁷.

The known form of ball-and-chain inactivation requires an intact S6 gating mechanism, but it is easy to imagine some interesting alternatives. For instance, an S6 gate that could not close all the way might not itself restrict ion movements, but might nevertheless regulate binding of the inhibitory peptide. This would produce a channel that would close whenever the S6 gate is in the open position. Alternatively, instead of using the S6 gate to regulate binding of the peptide, a channel protein might regulate the availability of the N terminus, or limit its access to the inner mouth of the channel. For instance, in one K^+ channel the inactivation process can be eliminated by experimentally induced disulphide formation between the N-terminal ball and another (unidentified) cysteine, effectively tying the ball out of the way²⁸. Channel sequences that are antagonistic to ball-and-chain inactivation²⁹ may function as decoy binding sites, luring the peptide away from its real site of action. β -subunits might regulate the availability of their N-terminal peptide through some action of their as-yet-unknown NAD-dependent enzymatic redox function³⁰. Finally, both α - and β -subunit N termini (as well as permeant ions) must approach the pore through 'windows' between the T1 domain and the transmembrane pore domain^{31,32} (see Box 1). Access to and through these windows might be restricted by conformational changes or by the binding of protein partners.

Gating at the selectivity filter

A third mechanism for closing the pore is to pinch shut at the

narrow selectivity filter itself—a selectivity filter or pore gate (Fig. 2, top right panel). This mechanism was first recognized in the form of ‘C-type inactivation’, an alternative to N-type inactivation for producing transient K^+ conductance by closing the channels in spite of a maintained stimulus. C-type inactivation persists in Shaker channels when the inactivation ball is deleted³³, and it is quite sensitive to extracellular $[K^+]$ and to mutations at the extracellular entryway to the pore³⁴. Closure of this gate can be prevented by the binding of extracellular tetraethylammonium ion (TEA) to the pore²⁵; once closed, this inactivation gate can be latched shut by a metal ion that bridges cysteines introduced at the outer mouth in each of the four subunits^{35,36}. This inactivation process (normally slower than N-type inactivation) probably has a function in regulating repetitive electrical activity; its sensitivity to $[K^+]$ may also determine a physiological response to accumulation

of extracellular K^+ (refs 37, 38).

The structural details of C-type inactivation probably resemble the deformation of the KcsA selectivity filter that is seen⁴ when those channels are crystallized with very low $[K^+]$: the selectivity filter contains fewer K^+ ions, and instead of pointing towards the central axis, the carbonyl oxygens of the filter project obliquely, with a partial collapse of the filter. In some voltage-gated K^+ channels, the inactivated selectivity filter can continue to conduct Na^+ ions when all K^+ is removed, suggesting a slightly different collapsed form where the carbonyl oxygens can now comfortably accommodate a Na^+ ion³⁹.

Single channel measurements^{40,41} in voltage-gated K^+ channels have suggested that even during the normal activation process (which at least in Shaker involves the S6 gate¹⁴), there may be movements of the selectivity filter. But gating at the selectivity filter may have an even more prominent role in other relatives of the voltage-gated K^+ channels. In CNG channels⁴² (but not HCN channels⁴³) the selectivity filter apparently acts as the principal gate of the channel: there is nearly free access from the intracellular solution to the S6 region even in the closed state of the channel, although there is some change in size restriction with gating⁴², and there are clear conformational changes in S6 (ref. 44).

Overall, it seems probable that the basic functioning of the S6 gate and the selectivity filter gate are conserved among different channels: the S6 gate moves in response to a transmembrane voltage sensor or intracellular sensor domain, and the selectivity filter may react to the change in S6 by opening or closing. The conserved glycine that acts as a gating hinge in the MthK channel probably helps to decouple these two motions, but not completely. Depending on the structural details, the S6 gate may close completely (as in the Shaker and HCN channels), or may remain ajar even in the ‘closed’ state (as in the CNG and small-conductance, Ca^{2+} -activated K^+ (SK) channels). Also depending on the details, the selectivity filter gate may remain always open, may respond to S6 movement by opening or closing, or may respond directly to voltage sensor motions.

Gating sensors for voltage and ligands

Control of these gating motions by voltage (or in the case of other family members, by ligand binding) is essential for the channels’ signalling function. Before considering control by voltage, I will discuss two sensor mechanisms that are better understood at the structural level. Both of these involve intracellular sensor domains that are directly attached to the C-terminal end of the S6 region (a motif common to many members of the family; Box 1).

The MthK channel—a bacterial 2TM channel already mentioned for its importance for understanding both permeation and gating—has recently given us a new insight into one possible sensor mechanism used by the tetrameric 6TM channels. Similar to other channels in the ligand-gated branch of the family (Box 1, lower right panel), the MthK channel has a C-terminal sensor domain. The RCK (regulator of K^+ conductance) domain used here is found also in other K^+ channels such as an *Escherichia coli* 6TM channel and in a mammalian high-conductance, Ca^{2+} -activated K^+ (BK) channel⁴⁵. The original finding that RCK domains form homodimers fit with the general impression that the four C-terminal sensor domains of a tetrameric channel might function as two dimers—a ‘dimer of dimers’ pattern⁴⁶. But the MthK channel was a surprise: a functional channel contains eight RCK domains, one on each channel subunit and one made in a soluble form by initiation at an internal methionine of the bacterial gene⁴⁷. A tight and invariant dimer interface connects each channel domain with a soluble domain, and these four dimers (each displayed as a single block in Fig. 3, top panel) then form a ring with four-fold symmetry. A second interface (between the neighbouring dimers) has a different appearance in the two known crystal forms of the RCK domains, suggesting a remarkable and

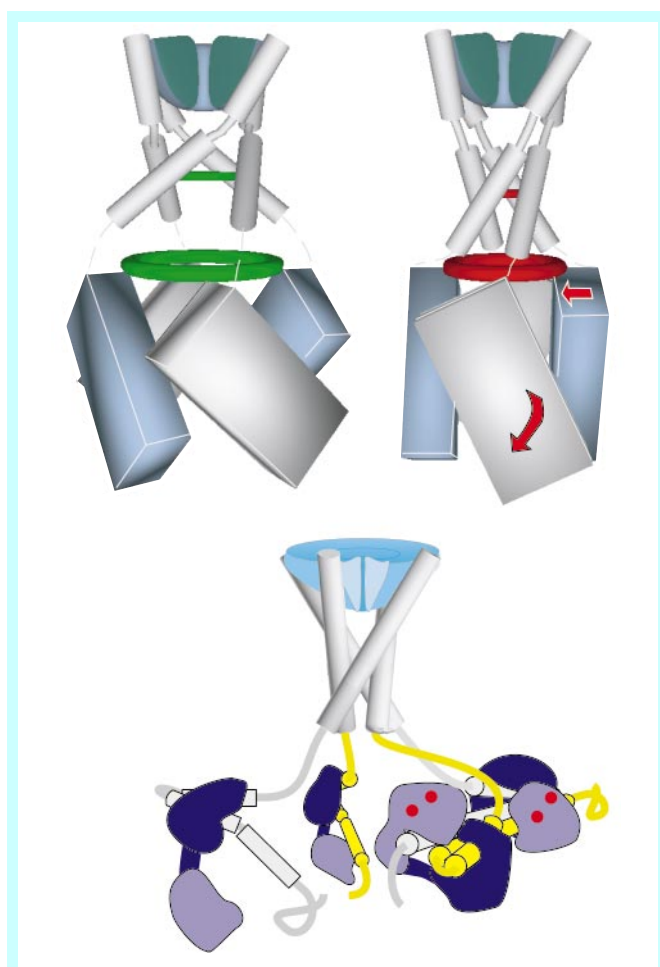


Figure 3 Two sensor domains that govern gating in the K^+ channel family. The top panel shows the proposed sensor-gating mechanism⁴⁷ for the calcium-gated MthK channel. Each rectangular box represents a dimer of two RCK domains (one attached to the channel subunit, one produced as a soluble protein). The open structure on the left has been observed for MthK with Ca^{2+} bound (although the linker between the channel domain and the sensor domain is not resolved). The closed structure on the right is proposed, on the basis of the closed channel structure of KcsA³ and the related RCK domain from the *E. coli* K^+ channel⁴⁵. A change at the interface between neighbouring RCK dimers (which would be produced in MthK by Ca^{2+} dissociation) could close the channel by rotating the dimers and contracting the ‘gating ring’. The bottom panel shows the proposed sensor mechanism for the calcium-gated SK channel⁶². The two subunits on the left show Ca^{2+} -free and monomeric CaM (blue). The two subunits on the right show how binding of Ca^{2+} ions (red) to CaM (blue) produces dimerization of the post-S6 C-terminal tails (shown alternately in grey and yellow).

elegant transduction mechanism. By changing from one form of the secondary interface to another, the 'gating ring' could contract and expand, prying open the S6 gate of the attached channel domain (Fig. 3, top panel).

Other ligand-gated channels in the family are likely to use multimeric sensor domains. The CNG channels use a C-terminal nucleotide-binding domain that is closely related to the dimeric catabolite activator protein of bacteria⁴⁸. The distantly related glutamate receptors—'upside down' K⁺ channels with their transmembrane domains flipped and the ligand sensor on the extracellular side⁴⁹—seem to have an important dimer relationship between their sensor domains⁵⁰.

Structural evidence for a possible dimer-of-dimers sensor mechanism comes from another type of Ca²⁺-activated channel that appears to signal by Ca²⁺-induced association of sensors connected to the S6 region^{51,52}. These 'SK' channels have a calmodulin-binding domain (CaMBD) located immediately adjacent to each S6; calmodulin (CaM) is stably bound here even in the absence of Ca²⁺. Studies of the detached CaM–CaMBD complex show that it behaves like a monomer in the absence of Ca²⁺, but dimerizes in the presence of Ca²⁺. The structure of this Ca²⁺-induced dimer shows an intimate reciprocal relationship between two CaM–CaMBD structures, with each CaM embracing both CaMBDs (Fig. 3, bottom). As in any dimer of dimers arrangement, formation of such dimers between neighbouring subunits would break the four-fold symmetry of the transmembrane channel; somehow this dimerization would produce a conformational change in the channel domain that opens the pore. Alternatively, with a modest change in the structure (and using the same protein interfaces seen in the crystal dimer structure), the four CaM–CaMBD complexes might form a four-fold symmetric tetramer upon Ca²⁺ binding.

The elusive voltage sensor

The voltage sensor is obviously central to the function of the voltage-gated K⁺ channels, but its three-dimensional structure is still unknown. There is, however, a great deal of functional information about voltage sensing.

The transmembrane voltage must be sensed by a structure that moves charge at least partially across the membrane. The energetics of voltage gating, together with direct electrical measurements of the gating charge movement, show that as many as four charges per channel subunit are effectively translocated across the membrane as the channel switches between closed and open states⁵³.

The expectation that transmembrane charges would be responsible for voltage sensing made it relatively easy to recognize the probable voltage sensor in the first channels cloned and sequenced⁵⁴. The only sequence motif conserved across all voltage-gated ion channels (Na⁺, Ca²⁺, K⁺) is in the fourth transmembrane region (S4), which at every third position has a positively charged arginine or lysine residue. Mutations of these charges have complicated effects on voltage gating, but the ultimate interpretation of these experiments is that the S4 charges are indeed the critical gating charges^{55,56}. These charges in the transmembrane region are energetically disfavoured (as discussed above for ion permeation); two possible mechanisms to compensate for this are negative countercharges in the S2 transmembrane region^{55,57} and water-filled 'canals'⁵⁸ at each end of the S4 transmembrane region. The canals also focus the voltage drop across a thin septum, allowing a small physical movement to have large electrostatic consequences⁵⁸.

Translocation of these S4 charges from one side of the membrane to the other can be confirmed biochemically, by replacing a charged residue with cysteine and applying cysteine modifying agents to one side or another^{59,60}. A position accessible from the intracellular side at negative voltage can become accessible from the outside at positive voltage. Remarkably little of the S4 is actually buried in the membrane: when one charge position is accessible from the

outside, a site only six residues away in the sequence is accessible from the inside. This too is consistent with the idea of water-filled access pathways at each end.

Altogether, the four segments S1–S4 are presumed to form the voltage sensor domain. On the basis of the pattern of sequence conservation, cysteine accessibility, and tolerance for substitution, all are probably mainly α -helical^{61–64}. The pattern of the charged residues at every third position led to the early proposal of a 'helical screw' motion^{65,66}; the S4 would advance screw-like to move each charge to the position of the next charge (or the next, and so on), thereby maintaining the charge distribution in the core while producing an overall translocation of charge across the membrane.

Many physical measurements have been made on fluorescence probes attached to cysteines placed in or near the S4, in an attempt to infer the motion of the voltage sensor. The pattern of fluorescence changes and the pattern of distances estimated from resonance energy transfer have roughly agreed with the idea of a rotation^{67,68}. Yet all of the measured distance changes (particularly from the measurements with the highest precision) are very small (about 1 Å) and seem inconsistent with the popular proposal of a substantial rotation. There are many other ways in which the effective translocation of charge could be accomplished, such as the relative lateral movement of crossed helices or a rocking motion at the interface between two domains. Solving this particular puzzle would be greatly helped by some knowledge of the three-dimensional structure.

The coupling problem

Unlike the ligand-binding domains attached directly to S6, the transmembrane voltage sensor does not have an obvious specific mechanism for coupling to the channel gates. Two possibilities have come to light recently. One is a direct coupling between the voltage sensor and the selectivity filter gate—in Shaker K⁺ channels, activation gating produces a close approach between the extracellular part of the S4 and the region flanking the selectivity filter⁶⁹. This then produces an allosteric effect on slow inactivation (involving the selectivity filter gate). At the intracellular side of the channel, a different site has been proposed⁷⁰ for coupling between S4 movement and the S6 activation gate: some sort of tight association between the S4–S5 linker and the intracellular projection of S6.

On the basis of the two best-known sensor mechanisms described in the previous section, it seems probable that a consistent theme of coupling will be that changes in quaternary structure of a sensor (such as ring expansion/contraction or domain multimerization) are transduced to the gating motions of the pore. It may therefore be tricky to identify a specific functional link between sensor and pore: the linkage may comprise a small number of 'tie-rods' with identifiable importance, or it may consist of an entire domain interface.

Dissecting these coupling mechanisms will of course be easier when we know the whole structure of a voltage-gated channel; however, particularly in this case, structure will not be enough, and careful mutagenesis and functional measurements will probably be required too. The fascination of the coupling problem is enhanced when one remembers that certain channels (for example, the HCN channels) have opposite relationships between voltage sensor movement and S6 gating: it appears that the same S6 gate used by 'normal' voltage-gated K⁺ channels closes at positive voltages rather than negative voltages⁴³ (although S4 has the normal positive charges and presumably the normal direction of motion).

Outlook

A combination of functional probing of channel proteins and recent structural advances have given clear pictures for many of the essential functions of the voltage-gated K⁺ channels and their relatives. The remarkable optimizations of these channels for permitting rapid and selective ion flow across the hostile barrier

of the cell membrane are now mostly apparent, as is the basic repertoire of conformational changes used to gate this flow. We now see the outlines of two general approaches used by intracellular sensor domains to manipulate the channel gates, and some tantalizing details of the transmembrane voltage sensor itself. Many challenges remain in learning how the intricate machinery is integrated into a functional whole, and into the larger function of cellular signalling. □

doi:10.1038/nature00978.

- Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)* **117**, 500–544 (1952).
- Latorre, R. & Miller, C. Conduction and selectivity in potassium channels. *J. Membr. Biol.* **71**, 11–30 (1983).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of potassium conduction and selectivity. *Science* **280**, 69–77 (1998).
- Zhou, Y., Morais Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion hydration and coordination revealed by a K⁺ channel-Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
- Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. Energetic optimization of ion conduction rate by the K⁺ selectivity filter. *Nature* **414**, 37–42 (2001).
- Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
- Roux, B. & MacKinnon, R. The cavity and pore helices in the KcsA K⁺ channel: electrostatic stabilization of monovalent cations. *Science* **285**, 100–102 (1999).
- Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. X-ray structure of a CIC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294 (2002).
- Armstrong, C. M. Ionic pores, gates, and gating currents. *Q. Rev. Biophys.* **7**, 179–210 (1975).
- Hodgkin, A. L. & Keynes, R. D. The potassium permeability of a giant nerve fibre. *J. Physiol. (Lond.)* **128**, 61–88 (1955).
- Hille, B. & Schwarz, W. Potassium channels as multi-ion single-file pores. *J. Gen. Physiol.* **72**, 409–442 (1978).
- Spassova, M. & Lu, Z. Coupled ion movement underlies rectification in an inward-rectifier K⁺ channel. *J. Gen. Physiol.* **112**, 211–221 (1998).
- Liu, Y., Holmgren, M., Jurman, M. E. & Yellen, G. Gated access to the pore of a voltage-dependent K⁺ channel. *Neuron* **19**, 175–184 (1997).
- del Camino, D. & Yellen, G. Tight steric closure at the intracellular activation gate of a voltage-gated K⁺ channel. *Neuron* **32**, 649–656 (2001).
- del Camino, D., Holmgren, M., Liu, Y. & Yellen, G. Blocker protection in the pore of a voltage-gated K⁺ channel and its structural implications. *Nature* **403**, 321–325 (2000).
- Holmgren, M., Shin, K. S. & Yellen, G. The activation gate of a voltage-gated K⁺ channel can be trapped in the open state by an intersubunit metal bridge. *Neuron* **21**, 617–621 (1998).
- Perozo, E., Cortes, D. M. & Cuello, L. G. Structural rearrangements underlying K⁺-channel activation gating. *Science* **285**, 73–78 (1999).
- Armstrong, C. M. Interaction of tetraethylammonium ion derivatives with the potassium channels of giant axon. *J. Gen. Physiol.* **58**, 413–437 (1971).
- Holmgren, M., Smith, P. L. & Yellen, G. Trapping of organic blockers by closing of voltage-dependent K⁺ channels: evidence for a trap door mechanism of activation gating. *J. Gen. Physiol.* **109**, 527–535 (1997).
- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* **250**, 533–538 (1990).
- Rettig, J. *et al.* Inactivation properties of voltage-gated K⁺ channels altered by presence of β-subunit. *Nature* **369**, 289–294 (1994).
- Zagotta, W. N., Hoshi, T. & Aldrich, R. W. Restoration of inactivation in mutants of Shaker K⁺ channels by a peptide derived from ShB. *Science* **250**, 568–571 (1990).
- Murrell-Lagnado, R. D. & Aldrich, R. W. Interactions of amino terminal domains of Shaker K channels with a pore blocking site studied with synthetic peptides. *J. Gen. Physiol.* **102**, 949–975 (1993).
- Murrell-Lagnado, R. D. & Aldrich, R. W. Energetics of Shaker K channels block by inactivation peptides. *J. Gen. Physiol.* **102**, 977–1003 (1993).
- Choi, K. L., Aldrich, R. W. & Yellen, G. Tetraethylammonium blockade distinguishes two inactivation mechanisms in voltage-activated K⁺ channels. *Proc. Natl Acad. Sci. USA* **88**, 5092–5095 (1991).
- Demo, S. D. & Yellen, G. The inactivation gate of the Shaker K⁺ channel behaves like an open-channel blocker. *Neuron* **7**, 743–753 (1991).
- Zhou, M., Morais-Cabral, J. H., Mann, S. & MacKinnon, R. Potassium channel receptor site for the inactivation gate and quaternary amine inhibitors. *Nature* **411**, 657–661 (2001).
- Ruppersberg, J. P. *et al.* Regulation of fast inactivation of cloned mammalian I_K(A) channels by cysteine oxidation. *Nature* **352**, 711–714 (1991).
- Roeper, J. *et al.* NIP domain prevents N-type inactivation in voltage-gated potassium channels. *Nature* **391**, 390–393 (1998).
- Gulbis, J. M., Mann, S. & MacKinnon, R. Structure of a voltage-dependent K⁺ channel beta subunit. *Cell* **97**, 943–952 (1999).
- Kobertz, W. R., Williams, C. & Miller, C. Hanging gondola structure of the T1 domain in a voltage-gated K⁺ channel. *Biochemistry* **39**, 10347–10352 (2000).
- Sokolova, O., Kolmakova-Partensky, L. & Grigorieff, N. Three-dimensional structure of a voltage-gated potassium channel at 2.5 nm resolution. *Structure (Camb.)* **9**, 215–220 (2001).
- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Two types of inactivation in Shaker K⁺ channels: effects of alterations in the carboxy-terminal region. *Neuron* **7**, 547–556 (1991).
- López-Barneo, J., Hoshi, T., Heinemann, S. H. & Aldrich, R. W. Effects of external cations and mutations in the pore region on C-type inactivation of Shaker potassium channels. *Recept. Channels* **1**, 61–71 (1993).
- Yellen, G., Sodickson, D., Chen, T.-Y. & Jurman, M. E. An engineered cysteine in the external mouth of a K⁺ channel allows inactivation to be modulated by metal binding. *Biophys. J.* **66**, 1068–1075 (1994).
- Liu, Y., Jurman, M. E. & Yellen, G. Dynamic rearrangement of the outer mouth of a K⁺ channel during gating. *Neuron* **16**, 859–867 (1996).
- Pardo, L. A. *et al.* Extracellular K⁺ specifically modulates a rat brain K⁺ channel. *Proc. Natl Acad. Sci. USA* **89**, 2466–2470 (1992).
- Baukowitz, T. & Yellen, G. Modulation of K⁺ current by frequency and external [K⁺]: a tale of two inactivation mechanisms. *Neuron* **15**, 951–960 (1995).
- Kiss, L., LoTurco, J. & Korn, S. J. Contribution of the selectivity filter to inactivation in potassium channels. *Biophys. J.* **76**, 253–263 (1999).
- Chapman, M. L., VanDongen, H. M. & VanDongen, A. M. Activation-dependent subconductance levels in the drk1 K⁺ channel suggest a subunit basis for ion permeation and gating. *Biophys. J.* **72**, 708–719 (1997).
- Zheng, J. & Sigworth, F. J. Selectivity changes during activation of mutant Shaker potassium channels. *J. Gen. Physiol.* **110**, 101–117 (1997).
- Flynn, G. E. & Zagotta, W. N. Conformational changes in S6 coupled to the opening of cyclic nucleotide-gated channels. *Neuron* **30**, 689–698 (2001).
- Rothberg, B. S., Shin, K. S., Phale, P. S. & Yellen, G. Voltage-controlled gating at the intracellular entrance to a hyperpolarization-activated cation channel. *J. Gen. Physiol.* **119**, 83–91 (2002).
- Johnson, J. P. Jr & Zagotta, W. N. Rotational movement during cyclic nucleotide-gated channel opening. *Nature* **412**, 917–921 (2001).
- Jiang, Y., Pico, A., Cadene, M., Chait, B. T. & MacKinnon, R. Structure of the RCK domain from the *E. coli* K⁺ channel and demonstration of its presence in the human BK channel. *Neuron* **29**, 593–601 (2001).
- Liu, D. T., Tibbs, G. R., Paoletti, P. & Siegelbaum, S. A. Constraining ligand-binding site stoichiometry suggests that a cyclic nucleotide-gated channel is composed of two functional dimers. *Neuron* **21**, 235–248 (1998).
- Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
- Weber, I. T. & Steitz, T. A. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. *J. Mol. Biol.* **198**, 311–326 (1987).
- Chen, G. Q., Cui, C., Mayer, M. L. & Gouaux, E. Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature* **402**, 817–821 (1999).
- Sun, Y. *et al.* Mechanism of glutamate receptor desensitization. *Nature* **417**, 245–253 (2002).
- Xia, X. M. *et al.* Mechanism of calcium gating in small-conductance calcium-activated potassium channels. *Nature* **395**, 503–507 (1998).
- Schumacher, M. A., Rivard, A. F., Bachinger, H. P. & Adelman, J. P. Structure of the gating domain of a Ca²⁺-activated K⁺ channel complexed with Ca²⁺/calmodulin. *Nature* **410**, 1120–1124 (2001).
- Sigworth, F. J. Voltage gating of ion channels. *Q. Rev. Biophys.* **27**, 1–40 (1994).
- Noda, M. *et al.* Primary structure of *Electrophorus electricus* sodium channel deduced from cDNA sequence. *Nature* **312**, 121–127 (1984).
- Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the Shaker K⁺ channel. *Neuron* **16**, 1159–1167 (1996).
- Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K⁺ channel. *Neuron* **16**, 1169–1177 (1996).
- Tiwari-Woodruff, S. K., Lin, M. A., Schulteis, C. T. & Papazian, D. M. Voltage-dependent structural interactions in the Shaker K⁺ channel. *J. Gen. Physiol.* **115**, 123–138 (2000).
- Islas, L. D. & Sigworth, F. J. Electrostatics and the gating pore of Shaker potassium channels. *J. Gen. Physiol.* **117**, 69–89 (2001).
- Yang, N., George, A. L. Jr & Horn, R. Molecular basis of charge movement in voltage-gated sodium channels. *Neuron* **16**, 113–122 (1996).
- Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. Transmembrane movement of the Shaker K⁺ channel S4. *Neuron* **16**, 387–397 (1996).
- Durell, S. R., Hao, Y. & Guy, H. R. Structural models of the transmembrane region of voltage-gated and other K⁺ channels in open, closed, and inactivated conformations. *J. Struct. Biol.* **121**, 263–284 (1998).
- Monks, S. A., Needleman, D. J. & Miller, C. Helical structure and packing orientation of the S2 segment in the Shaker K⁺ channel. *J. Gen. Physiol.* **113**, 415–423 (1999).
- Li-Smerin, Y., Hackos, D. H. & Swartz, K. J. Alpha-helical structural elements within the voltage-sensing domains of a K⁺ channel. *J. Gen. Physiol.* **115**, 33–50 (2000).
- Hong, K. H. & Miller, C. The lipid-protein interface of a Shaker K⁺ channel. *J. Gen. Physiol.* **115**, 51–58 (2000).
- Guy, H. R. & Seetharamulu, P. Molecular model of the action potential sodium channel. *Proc. Natl Acad. Sci. USA* **83**, 508–512 (1986).
- Catterall, W. A. Voltage-dependent gating of sodium channels: correlating structure and function. *Trends Neurosci.* **9**, 7–10 (1986).
- Glauner, K. S., Mannuzzu, L. M., Gandhi, C. S. & Isacoff, E. Y. Spectroscopic mapping of voltage sensor movement in the Shaker potassium channel. *Nature* **402**, 813–817 (1999).
- Cha, A., Snyder, G. E., Selvin, P. R. & Bezanilla, F. Atomic scale movement of the voltage-sensing region in a potassium channel measured via spectroscopy. *Nature* **402**, 809–813 (1999).
- Elinder, F., Mannikko, R. & Larsson, H. P. S4 charges move close to residues in the pore domain during activation in a K channel. *J. Gen. Physiol.* **118**, 1–10 (2001).
- Tristani-Firouzi, M., Chen, J. & Sanguinetti, M. C. Interactions between S4-S5 linker and S6 transmembrane domain modulate gating of HERG K⁺ channels. *J. Biol. Chem.* **277**, 18994–19000 (2002).
- Heginbotham, L., Abramson, T. & MacKinnon, R. A functional connection between the pores of distantly related ion channels as revealed by mutant K⁺ channels. *Science* **258**, 1152–1155 (1992).
- Ren, D. *et al.* A prokaryotic voltage-gated sodium channel. *Science* **294**, 2372–2375 (2001).
- Kreusch, A., Pfaffinger, P. J., Stevens, C. F. & Choe, S. Crystal structure of the tetramerization domain of the Shaker potassium channel. *Nature* **392**, 945–948 (1998).
- Gulbis, J. M., Zhou, M., Mann, S. & MacKinnon, R. Structure of the cytoplasmic beta subunit-T1 assembly of voltage-dependent K⁺ channels. *Science* **289**, 123–127 (2000).
- Shen, N. V. & Pfaffinger, P. J. Molecular recognition and assembly sequences involved in the subfamily-specific assembly of voltage-gated K⁺ channel subunit proteins. *Neuron* **14**, 625–633 (1995).

76. Holmes, T. C., Fadool, D. A., Ren, R. & Levitan, I. B. Association of Src tyrosine kinase with a human potassium channel mediated by SH3 domain. *Science* **274**, 2089–2091 (1996).
77. Doyle, D. A. *et al.* Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* **85**, 1067–1076 (1996).
78. Wallner, M., Meera, P. & Toro, L. Determinant for β -subunit regulation in high-conductance voltage-activated and Ca^{2+} -sensitive K^+ channels: an additional transmembrane region at the N terminus. *Proc. Natl Acad. Sci. USA* **93**, 14922–14927 (1996).
79. Morais, C. J. *et al.* Crystal structure and functional analysis of the HERG potassium channel N terminus: a eukaryotic PAS domain. *Cell* **95**, 649–655 (1998).
80. Knaus, H. G., Eberhart, A., Kaczorowski, G. J. & Garcia, M. L. Covalent attachment of charybdotoxin to the β -subunit of the high conductance Ca^{2+} -activated K^+ channel. Identification of the site of incorporation and implications for channel topology. *J. Biol. Chem.* **269**, 23336–23341 (1994).
81. Tapper, A. R. & George, A. L. Jr Location and orientation of minK within the I(Ks) potassium channel complex. *J. Biol. Chem.* **276**, 38249–38254 (2001).

Acknowledgements

I would like to thank members of my laboratory and B. Bean for many discussions and for suggestions about the manuscript, and R. MacKinnon for advice about timing. My research was supported by the National Institutes of Health.

Correspondence and requests for materials should be addressed to the author (e-mail: gary_yellen@hms.harvard.edu).

Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase

Javier Di Noia & Michael S. Neuberger

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

A functional immune system depends on the production of a wide range of immunoglobulin molecules. Immunoglobulin variable region (IgV) genes are diversified after gene rearrangement by hypermutation. In the DNA deamination model, we have proposed that deamination of dC residues to dU by activation-induced deaminase (AID) triggers this diversification. In hypermutating chicken DT40 B cells, most IgV mutations are dC → dG/dA or dG → dC/dT transversions, which are proposed to result from replication over sites of base loss produced by the excision activity of uracil-DNA glycosylase. Blocking the activity of uracil-DNA glycosylase should instead lead to replication over the dU lesion, resulting in dC → dT (and dG → dA) transitions. Here we show that expression in DT40 cells of a bacteriophage-encoded protein that inhibits uracil-DNA glycosylase shifts the pattern of IgV gene mutations from transversion dominance to transition dominance. This is good evidence that antibody diversification involves dC → dU deamination within the immunoglobulin locus itself.

The DNA deamination model¹ suggests that the initial event in IgV gene diversification is deamination of dC residues within the IgV gene by AID². The pathway employed for resolving the resultant dU/dG mismatch will then determine the precise nature of the diversification. Thus, replication templated on the dU/dG mismatch will lead to transitions at dC/dG (phase 1A mutation) (Fig. 1a). However, uracil residues in DNA are substrates for base excision repair³, and the action of a uracil-DNA glycosylase⁴ on the uracil will generate an abasic site (a site of base loss). Replication past this abasic site will result in dC → dA/dG/dT and dG → dA/dC/dT mutations. The precise pattern of nucleotide substitutions in this situation will probably depend on the preferences of the DNA polymerase involved but, given that two of the three possible substitutions are transversions, a transversion predominance is likely to ensue (phase 1B hypermutation). We have further envisaged that binding to the dU/dG mismatch (or to some intermediate in its repair) by the MSH2/MSH6 recognition complex could lead to error-prone patch repair, yielding mutations at dA/dT (phase 2 hypermutation), whereas recombination-mediated repair templated on an IgV pseudogene could lead to IgV gene conversion¹.

A critical test of this model would be provided by determining whether the hypermutation pathway could be perverted by inhibition of uracil-DNA glycosylase activity. In the case of phase 2 hypermutation and gene conversion, it is difficult to anticipate the effects of inhibiting uracil-DNA glycosylase because we do not know whether these pathways are triggered by the dU/dG mismatch itself or are triggered only after the generation of an abasic site. However, with regard to phase 1 hypermutation, it is clear that phase 1B mutation (templated on the abasic site) should be blocked by inhibition of uracil-DNA glycosylase activity, whereas phase 1A mutation (templated on the dU/dG itself) should be unaffected. Chicken DT40 B cells that lack the DNA-repair gene *XRCC2* (ref. 5) provide an attractive system in which to test this hypothesis because these cells constitutively perform transversion-favoured dC/dG (that is, largely phase 1B) hypermutation at high frequency, with little contribution of phase 2 mutation or gene conversion to IgV gene diversification⁶. More than 90% of the mutations in *XRCC2*-deficient DT40 cells are nucleotide substitutions at dC/dG, with some two-thirds of these substitutions being transversions⁶. The

different transversions are not equally favoured but rather suggest that the polymerase involved preferentially inserts dC opposite an abasic site.

Uracil-DNA glycosylase activity in DT40 cells

We were concerned that it might be difficult to obtain cells deficient in uracil-DNA glycosylase activity. Cytosine deamination is a major form of endogenous DNA damage⁷: inhibiting its repair by blocking all uracil-DNA glycosylase activity might be incompatible with clonal expansion. Gene redundancy might in any case make it difficult to generate glycosylase-deficient cells because there are several genes in the mammalian genome that encode enzymes able to excise uracil from DNA⁸. Indeed, mice carrying disruptions of the major uracil-DNA glycosylase gene (*UNG*, which is homologous to *E. coli ung*) exhibit a relatively mild phenotype⁹ and have revealed a backup uracil-excising activity that is probably encoded by the *SMUG1* DNA glycosylase gene^{10,11}.

We were struck, however, by the fact that we could not detect a *SMUG1* homologue in the well-represented chicken expressed sequence tag (EST) database (<http://chick.umist.ac.uk>), although ESTs homologous to human *UNG* were readily identified. We analysed the coding sequence of chicken *UNG*—obtained by polymerase chain reaction with reverse transcription (RT-PCR) from DT40 cells (see Supplementary Information)—and found high homology between the chicken and human enzymes. This suggested that it might be possible to inhibit the chicken *UNG* gene product using the uracil-DNA glycosylase inhibitor (Ugi) protein^{12,13} encoded by the uracil-containing genome of bacteriophage PSB-2. Ugi inhibits the *ung*-encoded uracil-DNA glycosylase from bacteria as well as from a wide variety of eukaryotic species including humans^{12–14}. The sequence of chicken *UNG* (see Supplementary Information) indicated that it retained the critical residues for catalytic activity and Ugi binding^{15,16}. This provided a way of ascertaining whether DT40 cells, like mouse cells¹¹, contain a major back-up activity for the excision of uracil from DNA when the *UNG* gene product is inhibited. No such back-up was apparent. Addition of Ugi to a DT40 cell extract before performing a uracil excision assay on an oligonucleotide substrate resulted in the inhibition of nearly all activity (Fig. 1b). Although this certainly

does not exclude the existence of a back-up uracil-excising activity in DT40 cells (such activity might simply not reveal itself under the particular *in vitro* assay conditions used), the result encouraged us to determine whether IgV gene diversification was affected in DT40 cells carrying a transfected Ugi expression cassette.

Inhibition of UNG in *Ugi* transfectants of DT40 cells

Electroporation of DT40 cells with a plasmid in which Ugi expression was linked to that of the selective marker by use of an internal ribosome entry site yielded very few transfectants (compared with controls), and those that survived the selection

contained no detectable *Ugi* messenger RNA as judged by northern blot analysis (Fig. 1c). Success was, however, achieved with a distinct vector (pEF-Ugi) in which Ugi expression was driven by the elongation factor 1 α (EF1 α) promoter and the selective marker (hygromycin resistance) was under control of the herpes simplex virus (HSV) thymidine kinase promoter (Fig. 1c). Of 30 clones analysed, all contained detectable *Ugi* mRNA (although at varying levels of abundance) as judged by northern blot. Extracts were prepared from 24 independent pEF-Ugi transfectants (12 from each of two electroporation experiments). All 24 extracts revealed a substantial (>99%) inhibition of uracil-DNA glycosylase

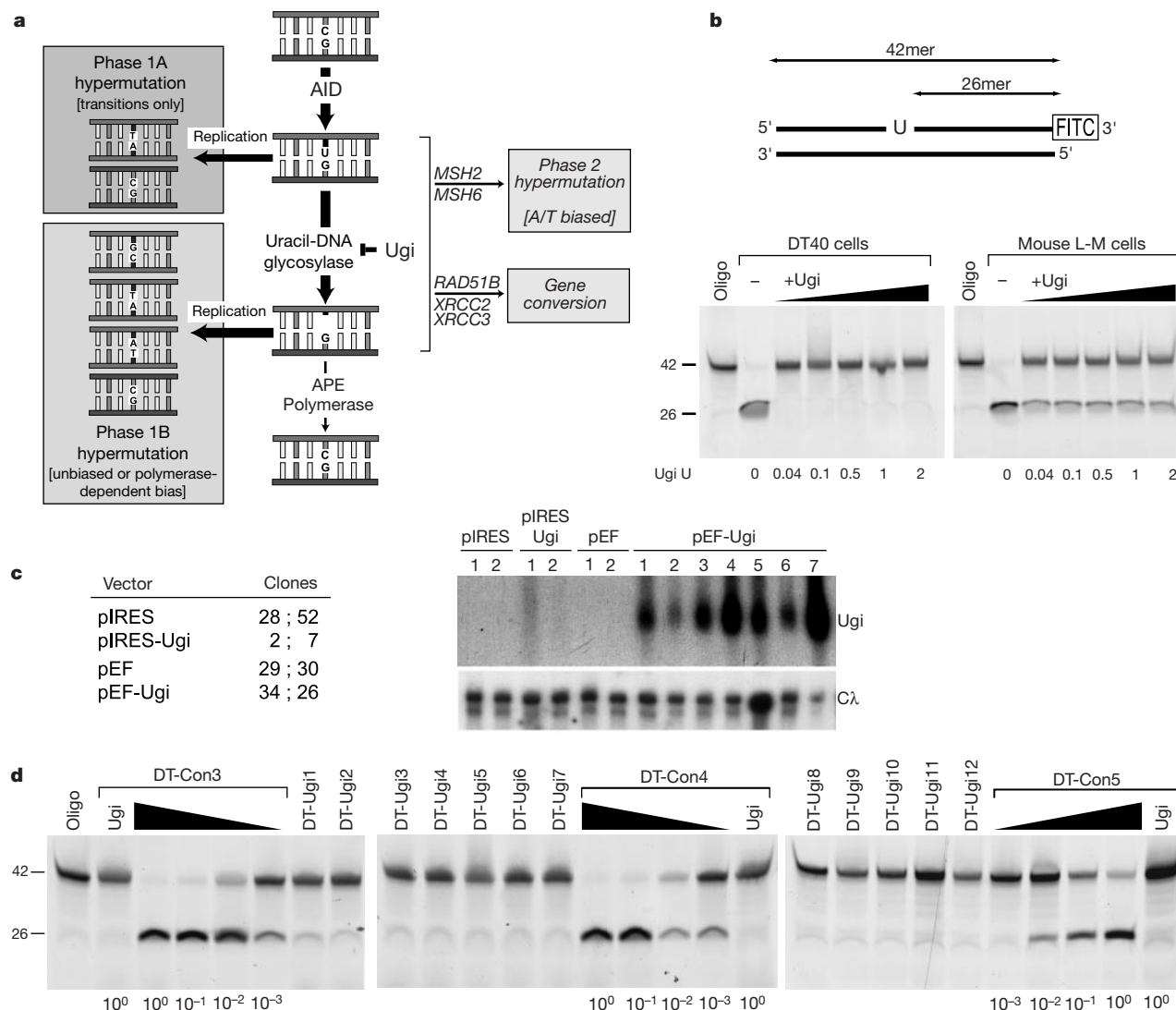


Figure 1 Uracil-DNA glycosylase activity in the DT40 B-cell line. **a**, The proposed role of uracil-DNA glycosylase in IgV gene diversification in *XRCC2*-deficient DT40 B cells. The major pathway of IgV gene diversification in *XRCC2*-deficient DT40 cells seems to be phase 1B hypermutation; phase 2 hypermutation and IgV gene conversion are minor contributors in these cells (see text) and are therefore indicated in italics. The extent to which the glycosylase-generated abasic site is repaired by an apyrimidic endonuclease (APE) and DNA polymerase (rather than simply acting as a template for replication—phase 1B mutation) is unknown, although work described here suggests that it is not the major pathway. **b**, Ugi inhibition reveals residual uracil-DNA excision activity in extracts of mouse L cells but not in DT40 cells. A fluorescein-labelled oligonucleotide containing a single dU residue was incubated with chicken DT40 or mouse L-M cell extracts (10 μ g protein) in the presence of varying concentrations of Ugi (0, 0.04, 0.1, 0.5, 1 and 2 units). The reaction products were resolved on 15% TBE-urea polyacrylamide gels.

c, Transfection of Ugi expression (and control) vectors into DT40 cells. The number of clones growing in selective medium obtained in two independent transfections performed with pIRES-Ugi and pEF-Ugi (as well as vector-only controls) is shown on the left. *Ugi* mRNA expression in randomly selected transfectants was analysed by northern blot, controlling with an immunoglobulin C λ probe. No *Ugi* mRNA was detected in any of the pIRES-Ugi transfectants analysed, whereas *Ugi* mRNA was detected (at varying levels) in all pEF-Ugi clones studied. **d**, Uracil-DNA excision activity in pEF-Ugi (and control) transfectants of *XRCC2*-deficient DT40 cells. Activity was monitored in serial dilutions of extracts from control transfectants (DT-Con3, 4 and 5; with 'Ugi' indicating that 2 units of purified Ugi protein were added to the extract at the time the assay was performed) or in undiluted extracts from pEF-Ugi transfectants (DT-Ugi1–12). Assays were performed as in **b** using undiluted extracts (10 μ g protein) as well as with tenfold serial dilutions (1:10, 1:100 and 1:1,000) in the case of DT-Con3, 4 and 5.

activity; no such inhibition was evident in 12 control transfectants (Fig. 1d).

Shifted IgV mutation pattern in *Ugi* transfectants

Somatic hypermutation in DT40 cells can lead to IgV gene inactivation, with such cells being readily identifiable by their loss of

surface immunoglobulin M (IgM). Comparison of *Ugi*-transfected *XRCC2*-deficient DT40 clones (DT-*Ugi* clones) and vector-only control transfectants (DT-Con clones) revealed no difference in the frequency of accumulation of surface IgM (sIgM)-loss variants: analysis of independent clones revealed a median of 5.1% sIgM-loss variants after 25 days in both cases. However, a substantial

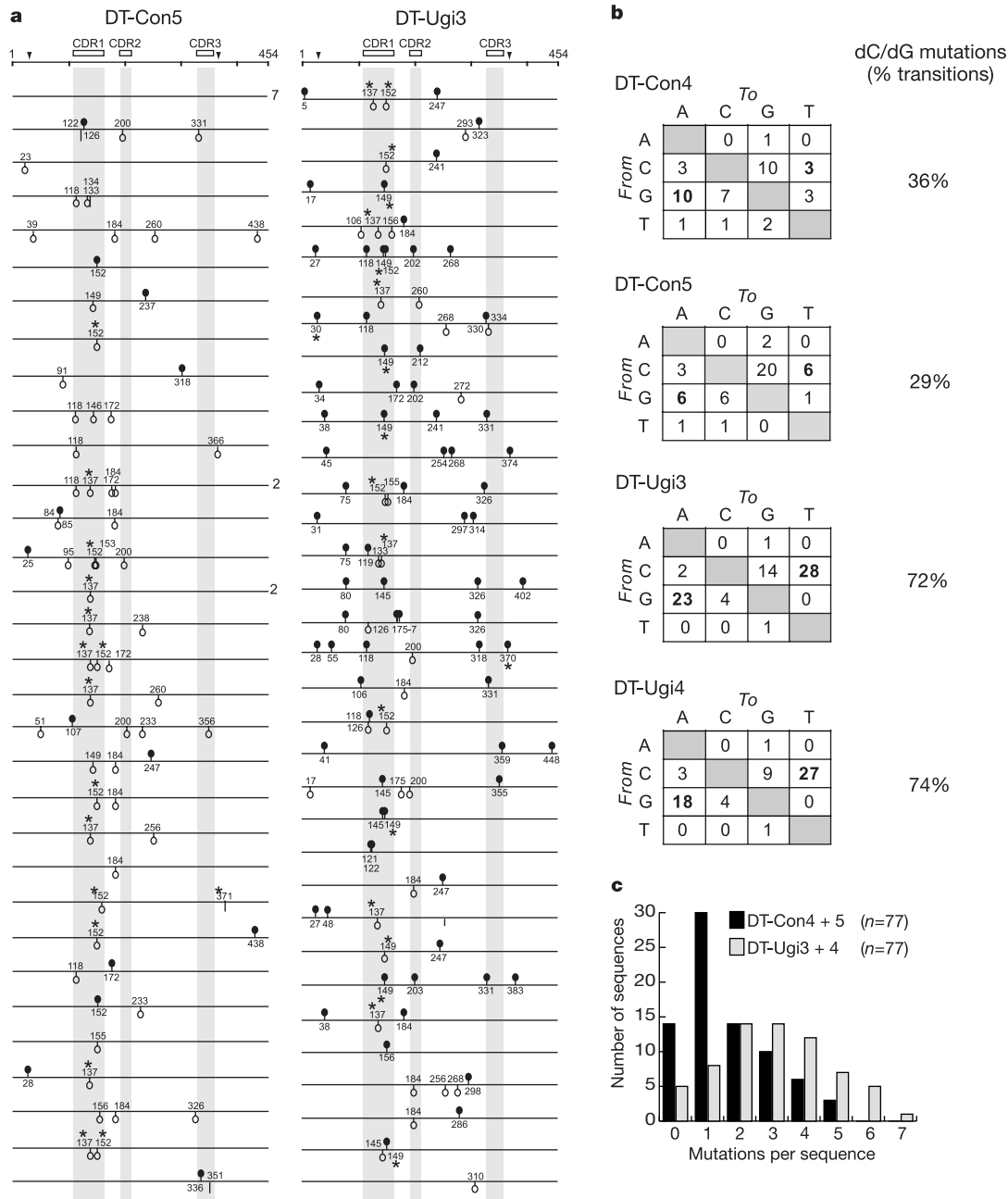


Figure 2 Analysis of Vλ mutations in surface IgM-loss variants sorted from pEF-*Ugi* and control transfectants. **a**, Comparison of Vλ sequences in sIgM-loss variants sorted from *Ugi*-expressing transfectant DT-*Ugi*3 and control transfectant DT-Con5 after 1 month of clonal expansion. Each horizontal bar represents the rearranged Vλ1/Jλ (numbering as in Fig. 3a). Transitions at dC/dG are indicated by a filled circle above the line, transversions at dC/dG by open circles below the line, and dA/dT mutations by a vertical line. Stop codons and splice site inactivations are indicated by an asterisk. A few Vλ clones of identical sequence were found among the population amplified from the sorted DT-Con5 (but not DT-*Ugi*3) loss variants; the number of such repeat sequences is indicated by a digit at the right end of the horizontal bar. No gene-conversion events (screened for by the criteria previously applied⁶) were identified in DT-*Ugi*3 or DT-Con5, although two

conversion events and a single base-pair deletion were identified among DT-*Ugi*4 sequences. **b**, Pattern of nucleotide substitutions. The digits within the tables indicate the number of the relevant substitution events observed. The same mutation when present in different Vλ sequences obtained from a single DT40 transfectant was scored only once. Thus, for example, the dC → dT transition at position 118 in the DT-*Ugi*3 clone was scored only once although it was observed in three distinct Vλ sequences. This stringent approach to avoiding overcalling mutations in clones that might be dynamically related will result in an undercalling of mutations, especially at mutational hotspots. **c**, Comparison of the mutational load in Vλ sequences obtained from *Ugi*-expressing (DT-*Ugi*) and control (DT-Con) clones.

difference was apparent between the two sets of cells when the Vλ sequences among the sorted sIgM-loss variants were compared (Fig. 2). Whereas nucleotide transitions accounted for only a minority of the substitutions at dC/dG in the control transfectants (13 of 36 and 12 of 42 in the two clones analysed), this increased to a majority (51 of 71 and 45 of 61) in the Ugi-expressing transfectants.

The mutation load was also somewhat greater in the two Ugi-expressing transfectants than in the controls (Fig. 2c), averaging 2.9 mutations per Vλ sequence for the sorted sIgM-loss variants from the Ugi-expressing clones, compared with 1.8 mutations per sequence in those obtained from sorted control cells. We suspect that this increased mutation load does not translate into a similar increase in the frequency of sIgM-loss variants because the transition-skewed nature of the base substitutions means that a smaller proportion of mutations lead to IgV gene inactivation. Thus, within the Vλ sequence, there are 17 dC or dG residues that can give rise to

a stop codon by a single nucleotide substitution but only 7 of these dC/dGs will yield a stop codon by a transition mutation. This cannot readily be extrapolated into calculating the effects of increased transition bias on the frequency of generation of sIgM-loss variants because not all dC/dGs are targeted equally and mutations other than stop codons in Vλ can lead to loss of sIgM expression. Nevertheless, the comparison is likely to reveal a general tendency.

To ensure that effects of Ugi expression on the mutation pattern were not obscured or skewed by selecting for sIgM-loss variants, we obtained IgVλ sequences from DT-Ugi3, DT-Ugi4, DT-Con4 and DT-Con5 as well as five other Ugi-expressing and five other control transfectants after 3–8 weeks of clonal expansion without imposing any selection for phenotypic changes. The mutations in all seven Ugi-expressing transfectants showed a marked skewing towards nucleotide transitions (Fig. 3 and Table 1). Although only 38% of

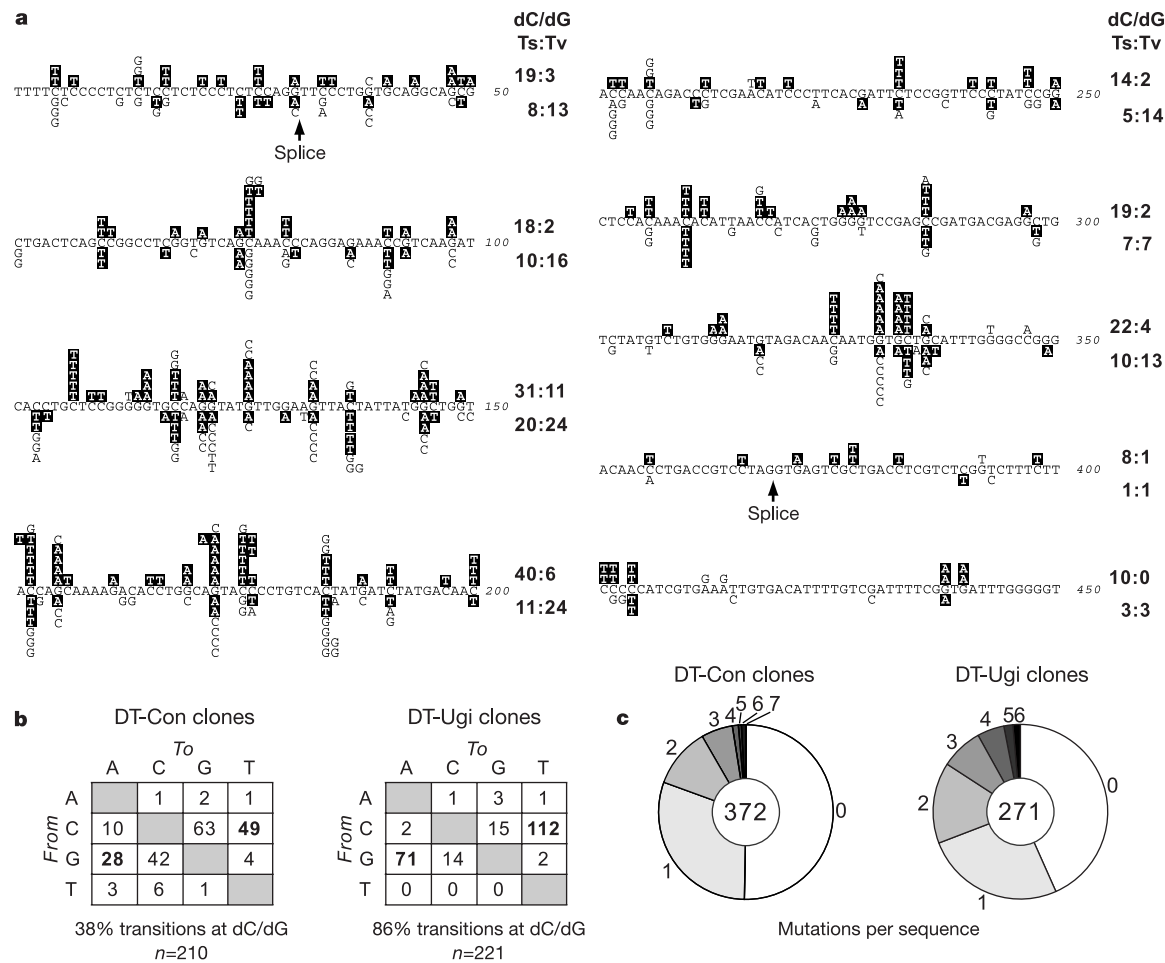


Figure 3 Analysis of Vλ mutations in unsorted populations of pEF-Ugi and control transfectants after 3–8 weeks of clonal expansion. **a**, Distribution of mutations across Vλ. The consensus Vλ1/Jλ sequence is the best fit to the Vλ sequences obtained from the 14 DT-Ugi and DT-Con clones. Mutations detected in the DT-Ugi-expressing clones are shown above the line, those in the control transfectants below the line, and transitions at dC/dG are shaded black. Adjacent to each line of Vλ sequence, the number of dC/dG transition (Ts) and transversion (Tv) mutations detected in that section of sequence in DT-Ugi (above the line) and DT-Con (below the line) transfectants is indicated. The same stringent criterion described in the legend to Fig. 2b was used to avoid repeat calling of the same mutational event. Only primary mutations are illustrated for reasons of clarity; 13 secondary mutations (6 in the DT-Ugi clones, 7 in the controls) are omitted from this panel and are designated secondary in that the parental clone in which the mutation arose already carried a mutation distinct from the consensus presented here at the relevant

position, or where dynastic relationships revealed that a mutated base had remutated during clonal expansion. These secondary mutations are, however, retained in the database underpinning the rest of the analyses. **b**, Pattern of nucleotide substitutions in the Ugi-expressing and control transfectants. The absolute numbers of mutations obtained from the individual Ugi-expressing and control clones (Table 1) have been separately compiled. The same stringent criterion applied in Fig. 2b has been used to avoid counting the same mutational event twice. **c**, Mutation load among the Vλ sequences from Ugi-expressing and control transfectants. The number in the centre of the pie chart is the number of Vλ sequences determined, with each slice of the pie indicating the proportion of sequences with 0–7 mutations. Each distinct Vλ sequence has been scored with no attempt made to remove dynastically related mutations from the calculation.

Table 1 λ mutations in Ugi-expressing and control DT40 transfectants

Clone	Expansion (weeks)	Sequences			Mutations at dC/dG		
		Number	Mutations per sequence	Total mutations	Ts	Tv	% Ts
DT-Ugi1	3	17	1.4	24	21	2	91
DT-Ugi2	3	22	1.1	24	19	5	79
DT-Ugi3	5	53	1.1	56	44	11	80
DT-Ugi4	5	58	0.7	38	31	7	82
DT-Ugi5	7	8	2.1	17	15	2	88
DT-Ugi6	7	30	0.6	17	14	3	82
DT-Ugi7	3	62	0.5	28	25	2	93
DT-Ugi7	8	21	1.3	28	23	2	92
Pooled DT-Ugi clones		271	1.1 $\pm$ 0.6	221	183	33	86 $\pm$ 6
DT-Con1	3	49	0.5	23	10	10	50
DT-Con2	3	30	0.8	24	4	20	17
DT-Con3	5	33	0.8	28	12	13	48
DT-Con4	5	45	0.6	27	10	15	40
DT-Con5	5	94	0.4	39	16	20	44
DT-Con6	5	43	0.6	26	12	12	50
DT-Con7	3	57	0.4	22	6	16	27
DT-Con7	8	21	1.2	26	7	18	28
Pooled DT-Con clones		372	0.6 $\pm$ 0.2	210	77	119	38 $\pm$ 13

Ts, transitions; Tv, transversions. In compiling the total number of mutations in each individual transfectant population, the same criterion used in Fig. 2b was applied to avoid counting the same mutational event twice. In calculating the total mutations in the pool of DT-Ugi and DT-Con clones, mutations common to the DT-Ugi7 (and DT-Con7) transfectants as analysed at 3 and 8 weeks have been counted only once. The average $\pm$ s.d. of mutations per sequence and % Ts are indicated for the pooled clones.

the nucleotide substitutions at dC/dG were transitions in the control transfectants (with the percentage in individual clones ranging from 17% to 50%), the percentage of transitions increased to a mean value of 86% in Ugi transfectants (79–93% in individual clones). It also seems that the Ugi-expressing clones accumulate slightly more mutations than control transfectants (despite the generation times of the Ugi-expressing transfectants averaging 2–3 h longer than controls; data not shown). Ugi expression, however, has had little effect on the distribution of mutations along the V gene segment, with 38% of the mutations occurring within an RGYW/WRCY nucleotide consensus (where R is purine, Y is pyrimidine and W is A or T) compared with 33% in controls⁶ (these percentages increase to 53% and 50%, respectively, if the consensus is restricted to RGY/RCY).

The effects of Ugi on the nature of the mutations accumulated within the IgV gene are likely to be through perversion of the pathway of antibody hypermutation in these cells, rather than a general mutagenic effect of inhibiting uracil-DNA glycosylase¹⁷. The mutations observed in Ugi-expressing cells are specific to the IgV gene segment: no mutations were found in either the Ig λ constant region or in a control gene. Thus, no mutations were detected in PCR-amplified genomic clones that covered either C λ (a stretch of 234 base pairs; 23 clones analysed) or a 540-base-pair portion of the AID gene (23 clones analysed). Similarly, introduction of Ugi into AID-deficient DT40 cells¹⁸ did not cause a mutator phenotype at the IgV locus because, like the parental AID-deficient DT40 cells, the Ugi-transfected AID-deficient cells did not give rise to sIgM-loss variants (the percentage of IgM-loss variants after 3 weeks was below the background detection level of 0.02% in both cases; data not shown).

Discussion

Inhibition of uracil-DNA glycosylase in XRCC2-deficient DT40 cells results in a small increase in the rate of mutation fixation within the V λ gene as well as a dramatic shift in the pattern of nucleotide substitutions at dC/dG from 38% to 86% transitions. The increased mutation rate could indicate that, in the absence of Ugi, a proportion of the dU residues in the IgV gene that have been generated by dC deamination are repaired: in the presence of Ugi, the repair pathway is now shifted into phase 1A mutation. It is equally possible that the increased mutation rate in the Ugi transfectants reflects the fact that whereas replication of the dU/dG pair (phase 1A) will always be mutagenic, replication over the abasic site (phase 1B) need not be mutagenic, depending on the

nucleotide inserted opposite the abasic site. At present, we cannot discriminate between these two possibilities.

However, the transition shift effected by inhibition of uracil-DNA glycosylase is readily explained by a shift from predominantly phase 1B mutation to predominantly phase 1A mutation. The reason why only 80–90% (rather than 100%) of the dC/dG mutations in Ugi-expressing cells are transitions is unclear. It could well be that not all uracil excision activity has been inhibited in these cells. Alternatively, even when uracil excision activity is inhibited, the cells might have available to them some other mutagenic pathway for processing or replicating past the dU in dU/dG lesions. In any case, the result gives a strong indication that IgV gene diversification does indeed proceed by a pathway involving the processing of dU residues generated by dC deamination within the Ig locus itself. $\square$

Methods

Cell culture, transfection and mutation analysis

DT40 subclone CL18 cells^{19,20}, its sIgM-positive derivative⁶ clone 4 as well as XRCC2-deficient⁵ and AID-deficient¹⁸ mutants were maintained as previously described⁶. Mouse L-Mtk⁺ fibroblasts were from our lab collection. The Ugi gene from *Bacillus subtilis* bacteriophage PBS2 (a gift from R. Savva) was cloned into the polylinkers of vector pIRESpu2 (Clontech) as well as into a derivative²¹ of pEAK8 (Edge Biosystems); an additional hygromycin-resistance cassette was then inserted into the BamHI site to yield pIRES-Ugi and pEF-Ugi. The plasmids were linearized (pIRES-Ugi, *PvuI*; pEF-Ugi, *MluI*) and transfected into XRCC2-deficient DT40 by electroporation^{6,18}; clones were selected in 1.5 mg ml⁻¹ hygromycin B (ICN) or 0.25 mg ml⁻¹ puromycin (Sigma).

The generation of sIgM-loss variants was analysed by flow cytometry and such variants were purified by cell sorting⁶. The rearranged V λ was amplified by PCR from genomic DNA using Pfu Turbo polymerase (Stratagene), cloned into M13mp18, and sequenced and analysed for mutation as previously⁶.

Uracil-DNA glycosylase assay

Exponentially growing cells were collected, washed in PBS buffer, resuspended in 25 mM Hepes, 5 mM EDTA, 1 mM dithiothreitol and 10% glycerol at pH 7.8 (HED buffer) with complete protease inhibitors (Roche) at about 10⁵ cells μ l⁻¹ and lysed by sonication (five 15-s pulses). After centrifugation at 10,000g for 20 min at 4 °C, the supernatant was frozen in aliquots into liquid N₂ and stored at -80 °C. Uracil-DNA glycosylase assays were carried out in HED buffer by mixing clarified whole-cell extract (10 μ g of total protein at the highest concentration) with 1 pmol of fluorescein-labelled oligonucleotide substrate in a final volume of 10 μ l for 2 h at 37 °C. This double-stranded oligonucleotide with a single dU/dG mismatch was made by annealing 5'-ATTATTATTATTCGUGGATTATTATTTTATTATTATTATTATTT-fluorescein to the complementary oligonucleotide, 5'-AAATAAATAAATAATAATAAATCCGCGGAATAATAATAAT. The reaction was terminated by the addition of 10 μ l of formamide loading dye (Amersham Pharmacia) and the products were resolved on 15% TBE-urea polyacrylamide gels. The reaction products were visualized using a Typhoon phosphorimager (Amersham Pharmacia). The existing AP endonuclease activity present in the extracts was sufficient to cleave all the deglycosylated substrate generated during the reaction. This was confirmed by performing parallel experiments in which the reaction was treated at the end of the incubation with *Escherichia coli* endonuclease IV (Trevigen) or alkali. In some experiments, the extract was pre-

incubated with purified Ugi (New England Biolabs) as indicated for 10 min on ice before adding the oligonucleotide substrate.

Received 28 June; accepted 16 July 2002; doi:10.1038/nature00981.

Published online 31 July 2002.

- Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99–103 (2002).
- Muramatsu, M. *et al.* Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**, 18470–18476 (1999).
- Lindahl, T. Suppression of spontaneous mutagenesis in human cells by DNA base excision-repair. *Mutat. Res.* **462**, 129–135 (2000).
- Lindahl, T. An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl Acad. Sci. USA* **71**, 3649–3653 (1974).
- Takata, M. *et al.* Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. *Mol. Cell. Biol.* **21**, 2858–2866 (2001).
- Sale, J. E., Calandrini, D. M., Takata, M., Takeda, S. & Neuberger, M. S. Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. *Nature* **412**, 921–926 (2001).
- Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
- Pearl, L. H. Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.* **460**, 165–181 (2000).
- Nilsen, H. *et al.* Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Mol. Cell* **5**, 1059–1065 (2000).
- Haushalter, K. A., Todd Stukenberg, M. W., Kirschner, M. W. & Verdine, G. L. Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr. Biol.* **9**, 174–185 (1999).
- Nilsen, H. *et al.* Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *EMBO J.* **20**, 4278–4286 (2001).
- Friedberg, E. C., Ganesan, A. K. & Minton, K. N-glycosidase activity in extracts of *Bacillus subtilis* and its inhibition after infection with bacteriophage PBS2. *J. Virol.* **16**, 315–321 (1975).
- Wang, Z. & Mosbaugh, D. W. Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in *Escherichia coli*. *J. Bacteriol.* **170**, 1082–1091 (1988).
- Karran, P., Cone, R. & Friedberg, E. C. Specificity of the bacteriophage PBS2 induced inhibitor of uracil-DNA glycosylase. *Biochemistry* **20**, 6092–6096 (1981).

- Mol, C. D. *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82**, 701–708 (1995).
- Handa, P., Roy, S. & Varshney, U. The role of leucine 191 of *Escherichia coli* uracil DNA glycosylase in the formation of a highly stable complex with the substrate mimic, Ugi, and in uracil excision from the synthetic substrates. *J. Biol. Chem.* **276**, 17324–17331 (2001).
- Radany, E. H. *et al.* Increased spontaneous mutation frequency in human cells expressing the phage PBS2-encoded inhibitor of uracil-DNA glycosylase. *Mutat. Res.* **461**, 41–58 (2000).
- Harris, R. S., Sale, J. E., Petersen-Mahrt, S. K. & Neuberger, M. S. AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Curr. Biol.* **12**, 435–438 (2002).
- Baba, T. W., Giroir, B. P. & Humphries, E. H. Cell lines derived from avian lymphomas exhibit two distinct phenotypes. *Virology* **144**, 139–151 (1985).
- Buerstedde, J. M. *et al.* Light chain gene conversion continues at high rate in an ALV-induced cell line. *EMBO J.* **9**, 921–927 (1990).
- Harris, R. S., Croom-Carter, D. S., Rickinson, A. B. & Neuberger, M. S. Epstein-Barr virus and the somatic hypermutation of immunoglobulin genes in Burkitt's lymphoma cells. *J. Virol.* **75**, 10488–10492 (2001).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank R. Savva for the Ugi cassette, S. Takeda for XRCC2-deficient DT40 cells, R. Grenfell for help with cell sorting, and R. Harris, S. Petersen-Mahrt, C. Rada and J. Sale for discussions. J.D.N. was supported by a César Milstein fellowship and the Fundación Antorchas, Argentina.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S.N. (e-mail: msn@mrc-lmb.cam.ac.uk).

Spin vector alignment of Koronis family asteroids

Stephen M. Slivan

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Rm 54-410, Cambridge, Massachusetts 02139, USA

Studies of asteroid families—groups of asteroids that formed from the fragmentation of larger bodies—are of broad interest to solar system researchers because they can provide insights into collisional processes, as well as the interior structures, strengths, and compositions of asteroids. It is generally accepted that members of the Koronis family were created by collisional disruption of a homogeneous parent body¹ and therefore share the same formation age and subsequent collisional history. The temporal variations in observed brightnesses of the Koronis family members (a consequence of their rotation) are, however, larger than expected². Preferential alignment of spin vectors had been proposed² as a possible explanation, but recent modelling³ predicted that family formation yields random spin vectors among the resulting fragments. Both hypotheses have been untested by observations. Here I show that the actual distribution of spin vectors among the largest members of the Koronis family falls within markedly nonrandom ‘spin clusters’. Reconciling models of family formation and evolution with the unexpected alignments of spin obliquities and correlations with spin rates presents a new challenge in understanding asteroid collisional processes.

Over 200 members of the Koronis family of asteroids have been identified⁴ among the thousands of objects in the main belt between Mars and Jupiter. Even the largest members of this family are too small directly to observe their shapes and orientations from Earth; instead, their spin states must be inferred from rotation light curves (Fig. 1). A substantial amount of observational data is needed to disentangle the spin vector information from the effects of shape and changing illumination geometry (Fig. 2). I used over 200 light curves spanning 46 yr (Table 1) to derive the spin vectors for nine of the largest members of the Koronis family. Most of the data were obtained between 1983 and 2001 during a series of observing programs that specifically targeted Koronis family asteroids^{5–8}.

Four different and independent methods of pole determination were used together to derive the spin vectors of the Koronis family sample objects. Two of the methods, weighted amplitude–aspect (WAA)⁹ and simultaneous amplitude–magnitude–aspect (SAM)⁹, are amplitude–magnitude methods which model an asteroid as a uniformly bright, featureless, smooth triaxial ellipsoid stably rotating about its shortest axis. WAA makes use of the dependence of amplitude on viewing aspect; SAM models both amplitudes and overall brightnesses. To the extent that an asteroid is well represented by the model, the two methods yield the orientation of the rotation axis, the axial ratios of the model ellipsoid, and the dependence of amplitude and magnitude on solar phase angle.

Amplitude–magnitude methods by themselves cannot distinguish between prograde and retrograde rotation because light curve amplitudes and magnitudes are unaffected by the sense of spin of the object. Epoch methods of pole determination complement amplitude–magnitude results by deducing the sense of spin and sidereal rotation period along with the spin axis orientation from small pole-dependent shifts in the observed times of light curve features. The epoch method that I used, sidereal photometric astrometry (SPA)⁹, assumes only that the chosen repeating ‘standard feature’ of the light curve corresponds to a fixed longitude on the asteroid.

The light curve data were also analysed using a convex polyhedron inversion approach^{10,11}. This method models the asteroid as a polyhedron with triangular faces, and unlike the amplitude–magnitude and epoch methods, it compares the corresponding model light curves with observed data points at all available rotational phases. It yields a model shape closely related to the convex hull of the body, as well as the sidereal rotation period, the sense of spin, and the orientation of the rotation pole.

The adopted pole solution for each object is the mean of its pole results from the separate methods, weighted by their estimated individual uncertainties. The overall estimated uncertainties for most of the pole solutions are about ten degrees of arc; the pole latitudes by themselves have estimated uncertainties of about five degrees. A detailed description of the different methods’ pole solutions for each object, including the individual results and errors, will be presented in a longer paper⁸.

Independent knowledge of the true pole of Koronis family member (243) Ida, determined from the Galileo spacecraft fly-by, was used to validate the analysis techniques. Ida’s sidereal period and a spin vector prediction had been determined from ground-based observations and accepted for publication before the encounter.

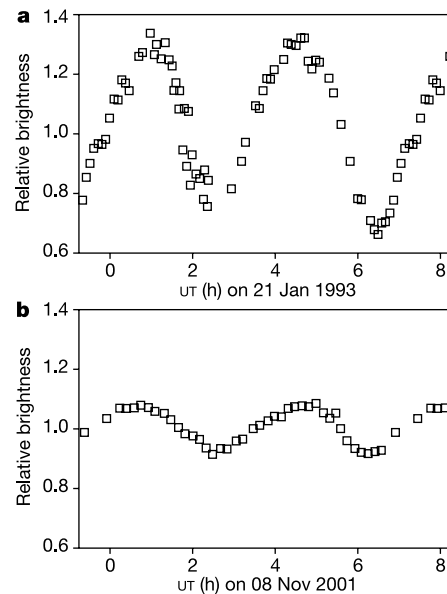


Figure 1 Light curves of (311) Claudia showing the doubly periodic change in brightness as it rotates once in 7.531 h. The brightness variations exhibited by a non-spheroidal asteroid as it rotates are caused mostly by the changes in the illuminated cross-sectional area presented to the observer. The amplitude of the resulting light curve thus depends upon both the object’s shape and the angle between the spin axis and the line of sight to the observer. Light curve observations at several different viewing geometries (**a**, **b**) are needed to determine an object’s spin vector direction. Koronis family asteroids have been observed specifically for spin vector determination since 1987 at five different observatories using 11 different telescope/detector combinations^{6–8}. Nearly 75% of these observations were made using the 0.61-m telescope at the George R. Wallace, Jr Astrophysical Observatory in Westford, Massachusetts, where a 384 × 576-pixel charge-coupled device (CCD) detector was used at Cassegrain focus with 5:1 focal length reduction for an image scale of 2.4 arcsec per pixel and a field size of 15′ × 23′. Integrations were usually 5 min using a V filter, and the telescope tracked unguided at the sidereal rate. The full-width at half-maximum of stellar images was typically 5–6″, and the synthetic aperture radius used for photometry was typically 12″. Asteroid brightnesses were measured relative to on-chip comparison stars used as secondary standards. On-chip stars of similar brightness to the asteroid were used to estimate the measurement uncertainties. All objects of interest were measured using the same aperture and sky annulus sizes. Relative brightnesses of comparison stars in different fields close together in the sky were measured within an hour or two of stable sky conditions on clear nights by imaging them rapidly enough to minimize differences in air mass and in time. UT, universal time.

Table 1 Summary of light curve observations for spin vector analysis

Object	Years observed	Apparitions	Nights	Observations
(158)	1984–1999	7	37	934
(167)	1977–1993	5	16	334
(208)	1985–1994	5	16	296
(277)	1983–2000	5	27	576
(311)	1984–2001	6	23	580
(321)	1955–2001	8	26	1,054
(534)	1977–1993	6	16	201
(720)	1983–1999	6	36	818
(1223)	1977–1996	7	18	738
Total:		55	215	5,531

An apparition is a period of several months during which the target object can be observed. Apparitions of any particular Koronis family asteroid occur at intervals of about 15 months.

ter¹²; subsequently its pole location was unambiguously derived from resolved images¹³. Applying the same spin vector determination methods used for the Koronis family sample objects to the available light curves of Ida gave a pole solution that is consistent with the true pole, subject to the limitations of the underlying simplifying assumptions made.

A twofold ambiguity is present for each sample object's pole solution: the two equally likely poles for each object have similar ecliptic latitudes and are about 180° apart in ecliptic longitude. It is not possible to identify which of the two solutions is the true pole because the low orbital inclinations of Koronis family members restrict Earth-based viewing geometries to be nearly symmetric with respect to the orbit plane. Nevertheless, the object's spin obliquity (that is, the angle between the spin pole and the orbit pole) can be determined in spite of the pole ambiguity because for low-inclination orbits the obliquity does not strongly depend on the pole longitude. For the Koronis sample objects the obliquities corre-

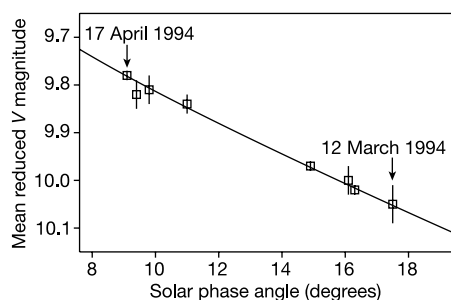


Figure 2 Effect of changing illumination geometry on the observed mean brightness of (158) Koronis during a 36-day interval in 1994. The horizontal axis is the angle between the line of sight and the direction of illumination from the Sun and the vertical axis is brightness, calibrated to standard *V* magnitudes and reduced to unit Earth–asteroid and Sun–asteroid distances. The solid line is the best-fit Lumme–Bowell solar phase function¹⁸. Solar phase effects must be removed from the calibrated observed light curves to combine data from different apparitions for spin axis determination. To calibrate observations to *V* magnitudes, standard stars¹⁹ with $V \approx 10$ –12 and $V-R$ colour ≈ 0.4 –0.6 were observed with the comparison stars during photometric conditions. Most of the total uncertainties in the overall transformation to standard *V* are 0.01–0.04 magnitude. Comparison stars used from 1998 to 2001 were calibrated by observing standards with the light curve observations. Comparison stars used from 1992 to 1994 were calibrated during four dedicated observing runs at two mountaintop sites^{6,7}. At the Michigan–Dartmouth–MIT Observatory 1.3-m telescope, CCD detectors binned to 512 × 512 pixels were used at Cassegrain focus for image scales of about 1'' per pixel and field sizes about 10' square. The telescope tracked autoguided at the sidereal rate, and the synthetic aperture radii for photometry were 5–6''. At the McDonald Observatory 0.91-m telescope, CCD detectors binned to about 260 × 260 pixels were used at Cassegrain focus for image scales of about 1'' per pixel and field sizes about 4' square. The telescope tracked unguided at the sidereal rate, and the synthetic aperture radius was 8''. Integrations were 15–30 s using a *V* filter. Stars that seemed variable or exhibited colour transformation effects were not used as comparison stars.

sponding to each pair of ambiguous poles differ by no more than 3°, and for most of the sample objects the difference is 1° or less, all well within the estimated uncertainties of the derived pole locations.

Including Ida, pole solutions are now available for ten of the largest members of the Koronis family. The histogram of the derived spin vector obliquities shown in Fig. 3a reveals a markedly bimodal distribution. I checked the statistical significance of the distribution using a Kolmogorov–Smirnov test¹⁴, a test for differences between cumulative distributions that is non-parametric (that is, no assumptions are made about the distributions of the populations) and that does not depend on the binning chosen for the data. Applying the test to the values of $\cos(\epsilon)$ shows that the distribution differs from a uniform distribution at greater than the 99% confidence level, indicating a statistically significant clustering of the spin vectors in obliquity. The results remain significant at the 95% confidence level even if pole solutions for five more family members were to yield completely random obliquities. Preliminary analysis of the light curve data available through 1994 had suggested the corresponding bimodal distribution in the spin axis latitudes⁶, but in the absence of analysis using an epoch method, confidence in the pole latitude results was low and the senses of rotation were unknown.

An unusual anisotropy is apparent in Fig. 3b, where correlation between the spin rates of the sample objects and their obliquities forms two distinct groupings, which I call 'spin clusters'. It is difficult to understand how preferred spin vector obliquities and rotation rates could exist in the context of formation by catastrophic fragmentation and disruption of a parent body, followed by at least 10⁹ yr of collisional evolution in the main belt¹⁵. Current models of

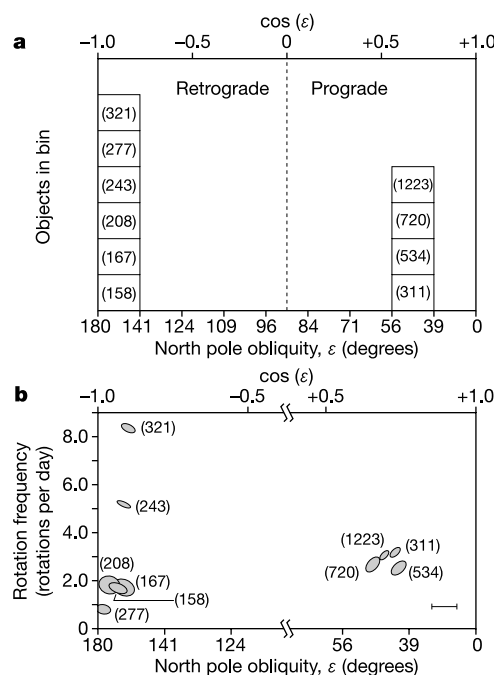


Figure 3 Spin angular momentum in the Koronis family of asteroids. **a**, Distribution of spin vector north pole obliquities ϵ for the sample objects, whose minor planet numbers are given in parentheses. Poles distributed isotropically on a sphere would give a uniform distribution in $\cos(\epsilon)$; the anisotropy of the sample obliquities is significant at greater than the 99% confidence level independent of the binning chosen. **b**, Spin vectors of Koronis family members plotted as rotation frequency versus $\cos(\epsilon)$, revealing two 'spin clusters' near (160°, 1.7) and (45°, 2.9). Each object is represented by an ellipse with size proportional to the object's diameter, axial ratio corresponding to the equatorial view of the object's model shape from its derived pole solution, and that has been rotated counterclockwise by the obliquity. The scale bar at lower right is 50 km. The empty central section of the plot is omitted to save space.

family formation predict that the memory of spin of the original unshattered parent body is lost³, and existing models of spin angular momentum suggest that collisional evolution randomizes asteroid spin vectors regardless of their initial orientations², although the absolute timescale is uncertain. Here, I briefly identify two possible general explanations for future study.

One possibility is that randomly oriented gravitational aggregates from the initial collision have further fragmented, creating smaller objects that have the same spin obliquities as the remnants from which they were formed. Secondary fragmentation of the largest remnant of the initial break-up has previously been proposed to explain the existence of several objects of comparable size among the largest Koronis family members¹⁶, but if the spin clusters were formed in this way then the absence of obvious corresponding associations in proper orbital elements also needs to be explained. To test this hypothesis, further work is needed to better understand the behaviour and evolution of gravitational aggregates.

A second possible explanation for spin clusters is that some dynamical process is aligning the obliquities and matching the rotation rates. Thermal effects can change obliquities and spin rates of small irregular asteroids, but calculations for Ida suggest that asteroids of comparable size are unlikely to have been substantially affected¹⁷. If a secular effect has clustered the spin vectors, then the present understanding of the timescale over which thermal processes have affected the spin cluster objects may be incomplete, or some nonthermal process may be at work. Finding similar clustering of spins for 20–40-km asteroids outside the Koronis family would support the hypothesis of a secular effect. □

Received 24 May; accepted 18 July 2002; doi:10.1038/nature00993.

- Chapman, C. R., Paolicchi, P., Zappala, V., Binzel, R. P. & Bell, J. F. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 386–415 (Univ. Arizona Press, Tucson, 1989).
- Binzel, R. P. Collisional evolution in the Eos and Koronis asteroid families: Observational and numerical results. *Icarus* **73**, 303–313 (1988).
- Michel, P., Benz, W., Tanga, P. & Richardson, D. C. Collisions and gravitational reaccumulation: Forming asteroid families and satellites. *Science* **294**, 1696–1700 (2001).
- Zappala, V., Bendjoya, Ph., Cellino, A., Farinella, P. & Froeschle, C. Asteroid families: Search of a 12,487-asteroid sample using two different clustering techniques. *Icarus* **116**, 291–314 (1995).
- Binzel, R. P. A photoelectric survey of 130 asteroids. *Icarus* **72**, 135–208 (1987).
- Slivan, S. M. *Spin-Axis Alignment of Koronis Family Asteroids* PhD thesis, Massachusetts Institute of Technology (1995).
- Slivan, S. M. & Binzel, R. P. Forty-eight new rotation lightcurves of 12 Koronis family asteroids. *Icarus* **124**, 452–470 (1996).
- Slivan, S. M. *et al.* Spin vectors in the Koronis family: Comprehensive results from two independent analyses of 213 rotation lightcurves. *Icarus* (submitted).
- Drummond, J. D., Weidenschilling, S. J., Chapman, C. R. & Davis, D. R. Photometric geodesy of main-belt asteroids. II. Analysis of lightcurves for poles, periods, and shapes. *Icarus* **76**, 19–77 (1988).
- Kaasalainen, M. & Torppa, J. Optimization methods for asteroid lightcurve inversion. I. Shape determination. *Icarus* **153**, 24–36 (2001).
- Kaasalainen, M., Torppa, J. & Muinonen, K. Optimization methods for asteroid lightcurve inversion. II. The complete inverse problem. *Icarus* **153**, 37–51 (2001).
- Binzel, R. P. *et al.* Asteroid 243 Ida: Ground-based photometry and a pre-Galileo physical model. *Icarus* **105**, 310–325 (1993).
- Davies, M. E. *et al.* The north pole direction and the control network of the asteroid 243 Ida. *Bull. Am. Astron. Soc.* **26**, 1154–1155 (1994).
- Miller, I. & Freund, J. E. *Probability and Statistics for Engineers* (Prentice-Hall, Englewood Cliffs, New Jersey, 1977).
- Belton, M. J. S. *et al.* First images of asteroid 243 Ida. *Science* **265**, 1543–1547 (1994).
- Marzari, E., Davis, D. & Vanzani, V. Collisional evolution of asteroid families. *Icarus* **113**, 168–187 (1995).
- Rubincam, D. P. Radiative spin-up and spin-down of small asteroids. *Icarus* **148**, 2–11 (2000).
- Bowell, E., *et al.* in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 524–556 (Univ. Arizona Press, Tucson, 1989).
- Landolt, A. U. *UVBRI* photometric standard stars around the celestial equator. *Astron. J.* **88**, 439–460 (1983).

Acknowledgements

I thank R. Binzel for advice and encouragement, and M. Kaasalainen, L. Crespo da Silva, M. Lyndaker and M. Krčo for contributions at various stages of this work.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to the author (e-mail: slivan@mit.edu).

Collapse and revival of the matter wave field of a Bose–Einstein condensate

Markus Greiner, Olaf Mandel, Theodor W. Hänsch & Immanuel Bloch

Sektion Physik, Ludwig-Maximilians-Universität, Schellingstrasse 4/III, D-80799 Munich, Germany, and Max-Planck-Institut für Quantenoptik, D-85748 Garching, Germany

A Bose–Einstein condensate represents the most ‘classical’ form of a matter wave, just as an optical laser emits the most classical form of an electromagnetic wave. Nevertheless, the matter wave field has a quantized structure owing to the granularity of the discrete underlying atoms. Although such a field is usually assumed to be intrinsically stable (apart from incoherent loss processes), this is no longer true when the condensate is in a coherent superposition of different atom number states^{1–6}. For example, in a Bose–Einstein condensate confined by a three-dimensional optical lattice, each potential well can be prepared in a coherent superposition of different atom number states, with constant relative phases between neighbouring lattice sites. It is then natural to ask how the individual matter wave fields and their relative phases evolve. Here we use such a set-up to investigate these questions experimentally, observing that the matter wave field of the Bose–Einstein condensate undergoes a periodic series of collapses and revivals; this behaviour is directly demonstrated in the dynamical evolution of the multiple matter wave interference pattern. We attribute the oscillations to the quantized structure of the matter wave field and the collisions between individual atoms.

In order to determine the evolution with time of a many-atom state with repulsive interactions in a confining potential, we first assume that all atoms occupy only the ground state of the external potential. The hamiltonian governing the system after subtracting the ground-state energy of the external potential is then solely determined by the interaction energy between the atoms:

$$\hat{H} = \frac{1}{2} U \hat{n}(\hat{n} - 1) \quad (1)$$

Here $\hat{n}$ counts the number of atoms in the confining potential, and U is the on-site interaction matrix element that characterizes the energy cost due to the repulsive interactions when a second atom is added to the potential well. It can be related to the s -wave scattering length a and the ground-state wavefunction $w(\mathbf{x})$ through $U = 4\pi\hbar^2 a/m \int |w(\mathbf{x})|^4 d^3x$, as long as the vibrational level spacing of the external potential is large compared with the interaction energy. The eigenstates of the above hamiltonian are Fock states $|n\rangle$ in the atom number, with eigenenergies $E_n = Un(n-1)/2$. The evolution with time (t) of such an n -particle state is then simply given by $|n\rangle(t) = |n\rangle(0) \times \exp(-iE_n t/\hbar)$, where $\hbar$ is Planck’s constant (h) divided by 2π .

We now consider a coherent state $|\alpha\rangle$ (see, for example, ref. 7) of the atomic matter field in a potential well. Such a coherent state with a complex amplitude α and an average number of atoms $\bar{n} = |\alpha|^2$ can be expressed as a superposition of different number states $|n\rangle$ such that $|\alpha\rangle = \exp(-|\alpha|^2/2) \sum_n \frac{\alpha^n}{\sqrt{n!}} |n\rangle$. Now the system is in a superposition of different eigenstates, which evolve in time according to their eigenenergies E_n . This allows us to calculate the evolution with time of an initially coherent state:

$$|\alpha\rangle(t) = e^{-|\alpha|^2/2} \sum_n \frac{\alpha^n}{\sqrt{n!}} e^{-iE_n(n-1)t/\hbar} |n\rangle \quad (2)$$

Evaluating the atomic field operator $\hat{a}$ for such a state then yields the

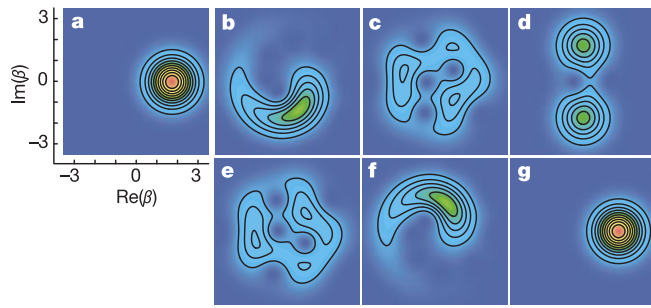


Figure 1 Quantum dynamics of a coherent state owing to cold collisions. The images **a–g** show the overlap $|\langle\beta|\alpha(t)\rangle|^2$ of an arbitrary coherent state $|\beta\rangle$ with complex amplitude β with the dynamically evolved quantum state $|\alpha(t)\rangle$ (see equation (2)) for an average number of $|\alpha|^2 = 3$ atoms at different times **a**, $t = 0\hbar/U$; **b**, $0.1\hbar/U$; **c**, $0.4\hbar/U$; **d**, $0.5\hbar/U$; **e**, $0.6\hbar/U$; **f**, $0.9\hbar/U$; and **g**, $\hbar/U$. Initially, the phase of the macroscopic matter wave field becomes more and more uncertain as time evolves (**b**), but remarkably at $t_{\text{rev}}/2$ (**d**), when the macroscopic field has collapsed such that $\psi \approx 0$, the system has evolved into an exact ‘Schrödinger cat’ state of two coherent states. These two states are 180° out of phase, and therefore lead to a vanishing macroscopic field ψ at these times. More generally, we can show that at certain rational fractions of the revival time t_{rev} , the system evolves into other exact superpositions of coherent states—for example, at $t_{\text{rev}}/4$, four coherent states, or at $t_{\text{rev}}/3$, three coherent states^{2,4}. A full revival of the initial coherent state is then reached at $t = \hbar/U$. In the graph, red denotes maximum overlap and blue vanishing overlap with 10 contour lines in between.

macroscopic matter wave field $\psi = \langle\alpha(t)|\hat{a}|\alpha(t)\rangle$, which has an intriguing dynamical evolution. At first, the different phase evolutions of the atom number states lead to a collapse of ψ . However, at integer multiples of the revival time $t_{\text{rev}} = \hbar/U$ all phase factors in the sum of equation (2) re-phase modulo 2π , leading to a perfect revival of the initial coherent state. The collapse time t_c depends on the variance σ_n^2 of the atom number distribution, such that $t_c \approx$

t_{rev}/σ_n (see refs 1–5). A more detailed picture of the dynamical evolution of ψ can be seen in Fig. 1, where the overlap of an arbitrary coherent state $|\beta\rangle$ with the state $|\alpha(t)\rangle$ is shown for different evolution times up to the first revival time of the many-body state^{8,9}.

In our experiment, we create coherent states of the matter wave field in a potential well, by loading a magnetically trapped Bose–Einstein condensate into a three-dimensional optical lattice potential. For low potential depths, where the tunnelling energy J is much larger than the on-site repulsive interaction energy U in a single well, each atom is spread out over all lattice sites. For the case of a homogeneous system with N atoms and M lattice sites, the many-body state can then be written in second quantization as a product of identical single-particle Bloch waves with zero quasi-momentum $|\Psi\rangle_{U/J=0} \propto (\sum_{i=1}^M \hat{a}_i^\dagger)^N |0\rangle$. It can be approximated by a product over single-site many-body states $|\phi_i\rangle$, such that $|\Psi\rangle_{U/J=0} \approx \prod_{i=1}^M |\phi_i\rangle$. In the limit of large N and M , the atom number distribution of $|\phi_i\rangle$ in each potential well is poissonian and almost identical to that of a coherent state. Furthermore, all the matter waves in different potential wells are phase coherent, with constant relative phases between lattice sites. As the lattice potential depth V_A is increased and J decreases, the atom number distribution in each potential well becomes markedly subpoissonian¹⁰ owing to the repulsive interactions between the atoms, even before entering the Mott insulating state^{11–13}. After preparing superposition states $|\phi_i\rangle$ in each potential well, we increase the lattice potential depth rapidly in order to create isolated potential wells. The hamiltonian of equation (1) then determines the dynamical evolution of each of these potential wells.

The experimental set-up used here to create Bose–Einstein condensates in the three-dimensional lattice potential (see Methods) is similar to that used in our previous work^{11,14,15}. Briefly, we start with a quasi-pure Bose–Einstein condensate of up to 2×10^5 ^{87}Rb atoms in the $|F=2, m_F=2\rangle$ state in a harmonic magnetic trapping potential with isotropic trapping frequencies of $\omega = 2\pi \times 24$ Hz. Here F and m_F denote the total angular momentum

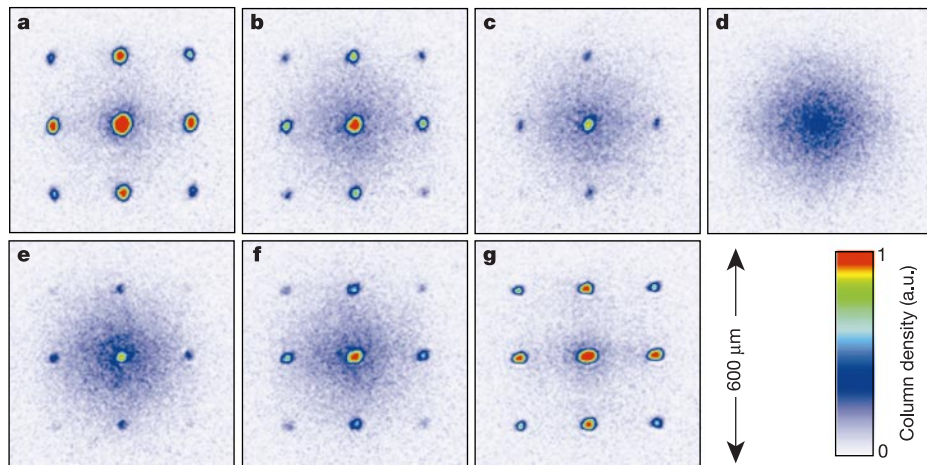


Figure 2 Dynamical evolution of the multiple matter wave interference pattern observed after jumping from a potential depth $V_A = 8 E_r$ to a potential depth $V_B = 22 E_r$ and a subsequent variable hold time t . After this hold time, all trapping potentials were shut off and absorption images were taken after a time-of-flight period of 16 ms. The hold times t were **a**, $0\mu\text{s}$; **b**, $100\mu\text{s}$; **c**, $150\mu\text{s}$; **d**, $250\mu\text{s}$; **e**, $350\mu\text{s}$; **f**, $400\mu\text{s}$; and **g**, $550\mu\text{s}$. At first, a distinct interference pattern is visible, showing that initially the system can be described by a macroscopic matter wave with phase coherence between individual potential wells. Then after a time of $\sim 250\mu\text{s}$ the interference pattern is completely lost. The vanishing of the interference pattern is caused by a collapse of the macroscopic matter wave field in each lattice potential well. But after a total hold time of $550\mu\text{s}$ (**g**) the interference pattern is almost perfectly restored, showing that the macroscopic matter

wave field has revived. The atom number statistics in each well, however, remains constant throughout the dynamical evolution time. This is fundamentally different from the vanishing of the interference pattern with no further dynamical evolution, which is observed in the quantum phase transition to a Mott insulator, where Fock states are formed in each potential well. From the above images the number of coherent atoms N_{coh} is determined by first fitting a broad two-dimensional gaussian function to the incoherent background of atoms. The fitting region for the incoherent atoms excludes $130\mu\text{m} \times 130\mu\text{m}$ squares around the interference peaks. Then the number of atoms in these squares is counted by a pixel-sum, from which the number of atoms in the incoherent gaussian background in these fields is subtracted to yield N_{coh} . a.u., arbitrary units.

and the magnetic quantum number of the atom's hyperfine state. In order to transfer the magnetically trapped atoms into the optical lattice potential, we slowly increase the intensity of the lattice laser beams over a time of 80 ms so that a lattice potential depth V_A of up to 11 recoil energies E_r (see Methods) is reached¹¹. This value of V_A is chosen so that the system is still completely in the superfluid regime¹⁶. We then rapidly increase the lattice potential depth to a value V_B of up to 35 E_r within a time of 50 μ s so that the tunnel coupling between neighbouring potential wells becomes negligible. The timescale for the jump in the potential depth is chosen such that it is fast compared with the tunnelling time between neighbouring potential wells, but sufficiently slow to ensure that all atoms remain in the vibrational ground state of each well. In this way, we preserve the atom number distribution of the potential depth V_A at the potential depth V_B .

We follow the dynamical evolution of the matter wave field after jumping to the potential depth V_B by holding the atoms in the optical lattice for different times t . After these hold times, we suddenly turn off the confining optical and magnetic trapping potentials and observe the resulting multiple matter wave interference pattern after a time-of-flight period of 16 ms. An example of such an evolution can be seen in Fig. 2, which shows the collapse and revival of the interference pattern over a time of 550 μ s. This collapse and revival of the interference pattern is directly related to the collapses and revivals of the individual coherent matter wave fields in each potential well. It is important to note a crucial difference between the outcome of a collapse and revival experiment in a double-well system and our multiple-well system. In a double-well system, a perfect interference pattern would be observed in each single realization of the experiment for all times. However, when the matter wave fields have collapsed in both wells, this interference pattern would alternate randomly for each realization. Averaging

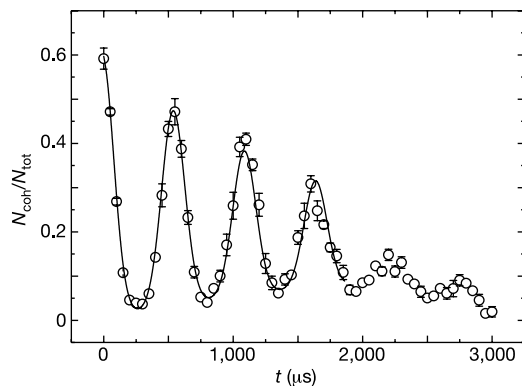


Figure 3 Number of coherent atoms relative to the total number of atoms monitored over time for the same experimental sequence as in Fig. 2. The solid line is a fit to the data assuming a sum of Gaussians with constant widths and constant time separations, including an exponential damping and a linear background. The damping is mainly due to the following process: after jumping to a potential depth V_B and thereby abruptly changing the external confinement and the on-site matrix element U , we obtain a parabolic profile of the chemical potential over the cloud of atoms in the optical lattice, which leads to a broadening of the interference peaks over time. When the interference peaks become broader than the rectangular area in which they are counted, we cannot determine N_{coh} correctly any more, which explains the rather abrupt damping that can be seen—for example, between the third and fourth revival in the above figure. Furthermore, the difference in U of $\sim 3\%$ over the cloud of atoms contributes to the damping of $N_{\text{coh}}/N_{\text{tot}}$ over time. The finite contrast in $N_{\text{coh}}/N_{\text{tot}}$ of initially 60% can be attributed to atoms in higher-order momentum peaks ($\sim 10\%$ of the total atom number), s -wave scattering spheres created during the expansion¹⁴, a quantum depletion of the condensate for the initial potential depth of $V_A = 8 E_r$, and a finite condensate fraction due to the finite temperature of the system.

over several single realizations would then yield the ensemble average value $\psi = 0$ that indicates the randomness of the interference pattern associated with the collapse of the matter wave fields. For the multiple-well set-up used here, however, the interference pattern in a single realization of the experiment can only be observed if the matter wave fields in each potential well have constant relative phase to each other, which requires that $\psi \neq 0$. The matter wave field ψ is therefore directly connected to the visibility of the multiple matter wave interference pattern in a single realization of the experiment.

In order to analyse quantitatively the temporal evolution of the interference pattern, we evaluate the number of atoms in the first and central order interference peaks N_{coh} versus the total number of atoms N_{tot} in the time-of-flight images. In the optical lattice, the matter wave field in each potential well $\psi_i(t) = \langle \phi_i(t) | \hat{a}_i | \phi_i(t) \rangle$ collapses and revives owing to the nonlinear dynamics discussed above. In order to relate the time evolution of the global fraction of coherent atoms $N_{\text{coh}}/N_{\text{tot}}$ to such a single-site time evolution $\psi_i(t)$ with $\bar{n}_i$ atoms on average on this lattice site, we sum the coherent fraction in each well over all M lattice sites: $N_{\text{coh}}/N_{\text{tot}} = 1/N_{\text{tot}} \sum_{i=1}^M |\psi_i(t)|^2$. This sum can be converted into an integral using the classical probability distribution $W(\bar{n})$ which describes the probability of finding a lattice site with an average number of $\bar{n}$ atoms. If the single-site dynamics is given by $\psi(t, \bar{n}, (U/J)_A, U_B)$, then the total number of coherent atoms can be determined by $N_{\text{coh}} = \int W(\bar{n}) |\psi(t, \bar{n}, (U/J)_A, U_B)|^2 d\bar{n}$. Using the Bose–Hubbard model and assuming a homogenous system, we are able to numerically calculate the initial atom number statistics on a single lattice site for finite U/J up to $U/J \approx 20$ and small $\bar{n}$ using a Gutzwiller ansatz^{13,17}. This allows us to predict the dynamical evolution of the matter wave field on a single lattice site $\psi(t, \bar{n}, (U/J)_A, U_B)$. Figure 3 shows the experimentally determined evolution of $N_{\text{coh}}/N_{\text{tot}}$ over time after jumping to the potential depth V_B . Up to five revivals are visible, after which a damping of the signal prevents further detection of revivals.

The revival of the matter wave field in each potential well is expected to occur at times that are multiples of $\hbar/U$, independent of the atom number statistics in each well. Therefore, in our inhomogeneous system, the macroscopic interference pattern should revive at the same times on all sites. As the on-site matrix element U increases for greater potential depths, we expect the revival time to decrease as V_B increases. This is shown in Fig. 4, where we have measured the revival period for different final potential depths V_B . We find excellent agreement between an *ab initio* calculation of $\hbar/U$

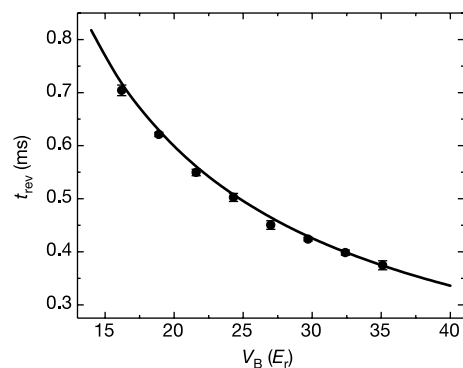


Figure 4 Revival period in the dynamical evolution of the interference pattern after jumping to different potential depths V_B from a potential depth of $V_A = 5.5 E_r$. The solid line is an *ab initio* calculation of $\hbar/U$ with no adjustable parameters based on a band structure calculation. In addition to the statistical uncertainties shown in the revival times, the experimental data points have a systematic uncertainty of 15% in the values for the potential depth.

from a band structure calculation and our data points. The revivals also directly prove the quantization of the underlying Bose field, and provide experimental proof that collisions between atoms lead to a fully coherent collisional phase $Un(n-1)t/2\hbar$ of the $|n\rangle$ -particle state over time, even on the level of individual pairs of atoms.

As we increase our initial lattice potential depth V_A , we expect the atom number distribution in each well to become markedly sub-poissonian owing to the increasing importance of the interactions as U/J increases. This in turn should lead to an increase of the collapse time, which depends on the variance of the superimposed number states. We have verified this by measuring the collapse time for different values of V_A (Fig. 5a). We can clearly observe a significant increase in the collapse time, when jumping from greater potential depths. For example, when jumping from $V_A = 11 E_r$, t_c/t_{rev} is more than 50% larger than when jumping from $V_A = 4 E_r$. This indicates that the atom number distribution in each potential well has indeed become subpoissonian, because for our experimental parameters the average atom number per lattice site, $\bar{n}_i$ remains almost constant when V_A is increased. A comparison of the collapse time for different initial potential depths V_A to a theoretical prediction is shown in Fig. 5b.

The observed collapse and revival of the macroscopic matter wave field of a Bose–Einstein condensate directly demonstrate behaviour of ultracold matter beyond mean-field theories. Furthermore, the

collapse times can serve as an independent, efficient probe of the atom number statistics in each potential well. It would be interesting to start from a Mott insulating state and use the coherent collisions between single atoms, which have been demonstrated here, to create a many-atom entangled state^{18–20}. This highly entangled state could then serve as a promising starting point for quantum computing with neutral atoms^{19,21}. □

Methods

Optical lattices

A three-dimensional array of microscopic potential wells is created by overlapping three orthogonal optical standing waves at the position of the Bose–Einstein condensate. The atoms are then trapped in the intensity maxima of the standing-wave light field owing to the resulting dipole force. The laser beams for the periodic potential are operated at a wavelength of $\lambda = 838$ nm with beam waists of $\sim 125 \mu\text{m}$ at the position of the Bose–Einstein condensate. This gaussian laser beam profile leads to an additional isotropic harmonic confinement of the atoms with trapping frequencies of 60 Hz for lattice potential depths of $20 E_r$. Here E_r denotes the recoil energy $E_r = \hbar^2 k^2 / 2m$, with $k = 2\pi/\lambda$ being the wavevector of the laser light and m the mass of a single atom. In this configuration, we populate almost 150,000 lattice sites with an average atom number per lattice site of up to 2.5 in the centre of the lattice. The lattice structure is of simple cubic type, with a lattice spacing of $\lambda/2$ and oscillation frequencies in each lattice potential well of ~ 30 kHz for a potential depth of $20 E_r$.

Received 9 April; accepted 4 July 2002; doi:10.1038/nature00968.

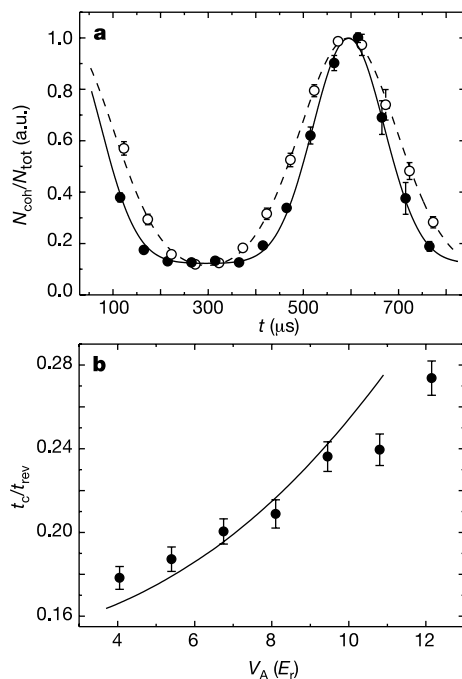


Figure 5 Influence of the atom number statistics on the collapse time. **a**, First revival observed in the ratio $N_{\text{coll}}/N_{\text{tot}}$ after jumping from different initial potential depths $V_A = 4 E_r$ (filled circles) and $V_A = 11 E_r$ (open circles) to a potential depth of $V_B = 20 E_r$. The data have been scaled to the same height in order to compare the widths of the collapse times, where the contrast of the curve at $V_A = 11 E_r$ was 20% smaller than that for $V_A = 4 E_r$. The solid and dashed line are fits to the data assuming a sum of two gaussians with constant widths t_c (measured as the $1/e$ half width of the gaussian), spaced by the corresponding revival time t_{rev} for the potential depth $V_B = 20 E_r$. **b**, Collapse time t_c relative to the revival time t_{rev} after jumping from different potential depths V_A to a potential depth $V_B = 20 E_r$. The solid line is an *ab initio* theoretical prediction based on the averaged time-evolution of the matter wave fields in each lattice potential well described in the text. Considering the systematic experimental uncertainties in the determination of the potential depths V_A of $\sim 15\%$ and an uncertainty in the total atom number of $\sim 20\%$, we find a reasonable agreement between both the experimental data and the theoretical prediction. a.u., arbitrary units.

1. Wright, E. M., Walls, D. F. & Garrison, J. C. Collapses and revivals of Bose-Einstein condensates formed in small atomic samples. *Phys. Rev. Lett.* **77**, 2158–2161 (1996).
2. Wright, E. M., Wong, T., Collett, M. J., Tan, S. M. & Walls, D. F. Collapses and revivals in the interference between two Bose-Einstein condensates formed in small atomic samples. *Phys. Rev. A* **56**, 591–602 (1997).
3. Imamoglu, A., Lewenstein, M. & You, L. Inhibition of coherence in trapped Bose-Einstein condensates. *Phys. Rev. Lett.* **78**, 2511–2514 (1997).
4. Castin, Y. & Dalibard, J. Relative phase of two Bose-Einstein condensates. *Phys. Rev. A* **55**, 4330–4337 (1997).
5. Dunningham, J. A., Collett, M. J. & Walls, D. F. Quantum state of a trapped Bose-Einstein condensate. *Phys. Lett. A* **245**, 49–54 (1998).
6. Zhang, W. & Walls, D. F. Bosonic-degeneracy-induced quantum correlation in a nonlinear atomic beam splitter. *Phys. Rev. A* **52**, 4696–4703 (1995).
7. Walls, D. F. & Milburn, G. J. *Quantum Optics* (Springer, Berlin, 1994).
8. Milburn, G. J. & Holmes, C. A. Dissipative quantum and classical Liouville mechanics of the anharmonic oscillator. *Phys. Rev. Lett.* **56**, 2237–2240 (1986).
9. Daniel, D. J. & Milburn, G. J. Destruction of quantum coherence in a nonlinear oscillator via attenuation and amplification. *Phys. Rev. A* **39**, 4628–4640 (1989).
10. Orzel, C., Tuchman, A. K., Fenselau, M. L., Yasuda, M. & Kasevich, M. A. Squeezed states in a Bose-Einstein condensate. *Science* **291**, 2386–2389 (2001).
11. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
12. Fisher, M. P. A., Weichman, P. B., Grinstein, G. & Fisher, D. S. Boson localization and the superfluid-insulator transition. *Phys. Rev. B* **40**, 546–570 (1989).
13. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
14. Greiner, M., Bloch, I., Mandel, O., Hänsch, T. W. & Esslinger, T. Exploring phase coherence in a 2D lattice of Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 160405–1–160405–4 (2001).
15. Greiner, M., Bloch, I., Hänsch, T. W. & Esslinger, T. Magnetic transport of trapped cold atoms over a large distance. *Phys. Rev. A* **63**, 031401–1–031401–4 (2001).
16. Cataliotti, F. S. et al. Josephson junction arrays with Bose-Einstein condensates. *Science* **293**, 843–846 (2001).
17. Rokhsar, D. S. & Kotliar, B. G. Gutzwiller projection for bosons. *Phys. Rev. B* **44**, 10328–10332 (1991).
18. Jaksch, D., Briegel, H. J., Cirac, J. I., Gardiner, C. W. & Zoller, P. Entanglement of atoms via cold controlled collisions. *Phys. Rev. Lett.* **82**, 1975–1978 (1999).
19. Briegel, H. J., Calarco, T., Jaksch, D., Cirac, J. I. & Zoller, P. Quantum computing with neutral atoms. *J. Mod. Opt.* **47**, 415–451 (2000).
20. Briegel, H. J. & Raussendorf, R. Persistent entanglement in arrays of interacting particles. *Phys. Rev. Lett.* **86**, 910–913 (2001).
21. Raussendorf, R. & Briegel, H. J. A one-way quantum computer. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).

Acknowledgements

We thank W. Zwerger, T. Esslinger, A. Görlitz, H. Briegel, E. Wright and I. Cirac for discussions, and A. Altmeyer for help in the final stages of the experiment. This work was supported by the DFG.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.B. (e-mail: imb@mpq.mpg.de).

The structure and chemistry of the TiO₂-rich surface of SrTiO₃ (001)

Natasha Erdman*, Kenneth R. Poeppelmeier†, Mark Asta‡, Oliver Warschkow§, Donald E. Ellis§ & Laurence D. Marks*

* Department of Materials Science and Engineering, Institute for Environmental Catalysis; † Department of Chemistry, Institute for Environmental Catalysis;

‡ Department of Materials Science and Engineering; and § Department of Physics and Astronomy, Institute for Environmental Catalysis, Northwestern University, Evanston, Illinois 60208-3108, USA

Oxide surfaces are important for applications in catalysis and thin film growth. An important frontier in solid-state inorganic chemistry is the prediction of the surface structure of an oxide. Comparatively little is known about atomic arrangements at oxide surfaces at present, and there has been considerable discussion concerning the forces that control such arrangements. For instance, one model suggests that the dominant factor is a reduction of Coulomb forces¹; another favours minimization of 'dangling bonds' by charge transfer to states below the Fermi energy². The surface structure and properties of SrTiO₃—a standard model for oxides with a perovskite structure—have been studied extensively^{3–14}. Here we report a solution of the 2 × 1 SrTiO₃ (001) surface structure obtained through a combination of high-resolution electron microscopy and theoretical direct methods. Our results indicate that surface rearrangement of TiO_{6-x} units into edge-sharing blocks determines the SrO-deficient surface structure of SrTiO₃. We suggest that this structural concept can be extended to perovskite surfaces in general.

SrTiO₃ has been used as a substrate for epitaxial growth of superconducting thin films¹³ and a buffer layer for the growth of GaAs thin films on Si¹⁴. The structure of SrTiO₃ can be described as a close-packed lattice of oxygen and strontium in a 3:1 ratio, with titanium in octahedral interstitial sites. For an ideal bulk-terminated (001) surface, two configurations are possible: (1) Ti and O termination with a ratio of 1:2 (TiO₂ stoichiometry); or (2) Sr and O termination with a ratio of 1:1 (essentially SrO). However, under various conditions (of atmosphere, pressure and temperature) transition metal oxides can exhibit different valences and coordination numbers, so the structure is actually more complex. Depending upon the experimental environment, under oxidative conditions 2 × 1, 4 × 2 and 6 × 2 surface reconstructions of SrTiO₃ have been reported^{4–8} together with more complicated structures obtained by annealing in vacuum under reducing conditions^{9,10}. In many cases the results are inconclusive and contradictory, and only a few studies have attempted to generate a model for these surface structures.

Diffraction techniques provide a powerful tool for the solution of surface structures and refinement of atomic positions. In a diffraction experiment, the amplitudes of the reflections are recorded, but the phase information is lost, preventing a direct Fourier inversion of the data. Our approach^{15–18} solves the phase problem by exploiting probability relationships between the amplitudes and the phases of the diffracted beams. The algorithm searches for the set of phases with the lowest figures of merit (FOM), which are then used to create scattering potential maps that obey the imposed symmetry. Effectively, a set of plausible solutions for the structure is generated, starting only from the intensity data. If the experimental errors are very small, these maps can approach an accurate restoration. For surface data, provided that the sample is tilted off the zone axis, the intensities are more than 90% kinematical¹⁹ and high-quality maps can generally be obtained. The method has already been used with either surface electron or X-ray diffraction data to solve more than

20 different surface structures using experimental data, more than half of which had not previously been solved. There are some subtle issues in using this approach to solve for surface structures, which are described in more detail elsewhere^{15,16}.

Starting with a single-crystal SrTiO₃ (001) wafer (99.95% pure), 3-mm disks were cut using an ultrasonic cutter. These were mechanically polished to a thickness of about 120 μm, dimpled, then ion milled with 4.8-kV Ar⁺ ions using a Gatan precision ion polishing system (PIPS) to produce an electron-transparent sample. The sample was annealed in a tube furnace at 950–1,000 °C with a constant flow of high-purity oxygen to eliminate the damage caused by ion milling and to obtain the 2 × 1 reconstruction. Annealing at lower or higher temperatures produces the 4 × 2 and 6 × 2 reconstructions respectively, which we do not discuss here. These results were highly reproducible, indicating that the 2 × 1 structure observed is thermodynamically stable. Off-zone-axis electron diffraction patterns and bright-field images were obtained using the ultrahigh vacuum (UHV)-H9000 Hitachi electron microscope²⁰, operated at 300 kV at Northwestern University. For high-resolution electron microscopy (HREM) the samples were tilted to the zone axis and images were recorded using a JEOL-4010 electron microscope operated at 400 kV at Argonne National Laboratory.

Figure 1 shows an off-zone-axis selected area diffraction pattern with both 2 × 1 and 1 × 2 domains (marked on the figure) combined with a bright-field image of the SrTiO₃ (001) sample. The bright-field image indicates the formation of large flat <100> facets at the surface of the sample after annealing. Both high-resolution images and the theoretical direct methods yielded the same general solution to the surface structure; for clarity we describe the images first. Figure 2a shows a high-resolution image of a mainly single-domain 2 × 1 surface and a corresponding power spectrum. The image was noise-filtered using a Wiener filter²¹ and the bulk contribution suppressed, similar to the strategies used previously^{22,23}. Figure 2b displays the same image after lattice averaging (left) and after filtering to separate translation/inversion domains of the structure (right) on the top and bottom surfaces (for more details see ref. 23) with an image simulation inset in the centre. The images show a dominant structural motif, a zig-zag row of features with a local *p2mg* symmetry. Numerical image simulations (inset in Fig. 2b) were in fairly good agreement with the experimental data, with black dots corresponding to surface titanium sites. Among the plausible solutions determined by direct methods, only those

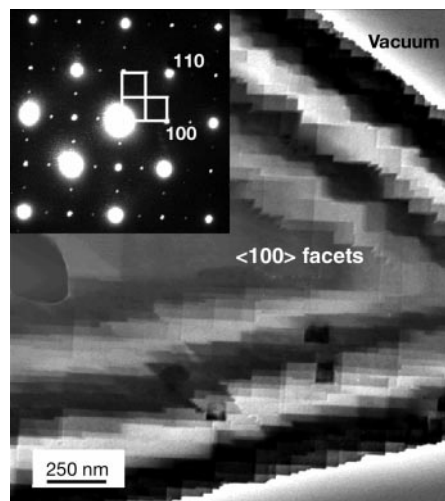


Figure 1 Bright-field image and off-zone-axis diffraction pattern (inset) of the (2 × 1) SrTiO₃ (001) surface, taken using the UHV-H9000 Hitachi transmission electron microscope (TEM). The surface unit cell for both domains of the reconstruction is marked.

containing the same structural element (for example, Fig. 3) were stable to subsequent refinements. Analysis indicated that these were Ti atom sites (not O or Sr), and using conventional difference maps combined with χ^2 refinements we were able to determine the oxygen sites.

The refinement of the structure against the diffraction data gave both the Ti and O atom positions, with a $\chi^2 = 2.01$. Alternative models of the structure, for example O addition on top of the Ti atoms, were considered, but the refinement results were significantly inferior.

The two-dimensional nature of the diffraction data does not allow refinement of the z position, so plane-wave pseudo-potential density functional (DFT) calculations^{24–27} on a 13-layer slab model were employed as an independent structure refinement in three dimensions, starting from the experimentally obtained atomic positions. The calculation confirmed the x,y positions obtained from the χ^2 refinement, as well as relaxing the structure in the z direction (see Supplementary Information for atom positions and partial charges). Figure 4 shows the solution of the 2×1 SrTiO₃ surface structure. The top layer has a TiO₂ stoichiometry, with half of the TiO₅ units displaced along the $[110]$ direction. The result is a row of 5-coordinated TiO₅ units that share edges between them as well as with the layer underneath, which is a part of the first full intact TiO₂ layer of SrTiO₃ with 6-coordinated Ti atoms. We note

the tilting of the TiO₆ octahedra in the bulk layers underneath to support the reconstruction. In the surface layer we can distinguish between two different types of O atoms: (1) ‘internal’ O atoms (O1, O2) tightly bound inside the surface rows and shared between the octahedra; and (2) ‘external’ O atoms (O3, O4), not shared between the octahedra. The distances between the surface Ti atoms to the two external O atoms are significantly shorter than the bulk Ti–O distances (1.51 Å and 1.76 Å versus 1.95 Å), whereas the distance between the internal oxygens to surface Ti is bulk-like. Apparently the surface Ti atoms compensate for the lower coordination number by stronger bonding to the external oxygens. We further employed near-minimal basis linear combination of atomic orbitals (LCAO) DFT calculations²⁸ to obtain partial atomic charges. In support of the above argument, the charge analysis finds that the surface Ti and internal O atoms are bulk-like (Ti +2.0, O –1.2), whereas the charges on the two external O atoms are significantly reduced (O3 –0.8 and O4 –0.6 compared with –1.2 in the bulk), which is indicative of increased covalence.

SrTiO₃ is representative of a large class of oxides with the perovskite structure; it is therefore important to understand the basic character of its surfaces because similar considerations will apply to other materials. It is well established that the coordination chemistry of perovskite structures is dominated by the formation of corner-shared octahedral building blocks, along with large cations to achieve the appropriate stoichiometry. In an ABO₃ structure the A cation prevents any structural rearrangement at the surface. Without these large cations, the ReO₃-type structure is created, which can have corner as well as edge-shared octahedra; this means additional degrees of freedom for structural rearrangements. On chemical grounds, we therefore expect that the surface of a perovskite will rearrange to a BO_{3–x} configuration, effectively screening the B cations, while also minimizing surface dipoles and maximizing the coordination.

The surface structure reflects these concepts and is a hybrid of three components: (1) The first component is a linear structure involving two Ti atoms (Ti1 and Ti2) and two O atoms (O1 and O2), with essentially bulk-like electronic structure similar to edge-shared octahedral units, where each O atom is coordinated to 3 or 4 Ti atoms. (2) The second component consists of two external O atoms (O3 and O4) with much more covalent character, particularly O4 which is double-bonded to Ti1; the coordination number for each of the external O atoms is 1 or 2. (3) The third component is

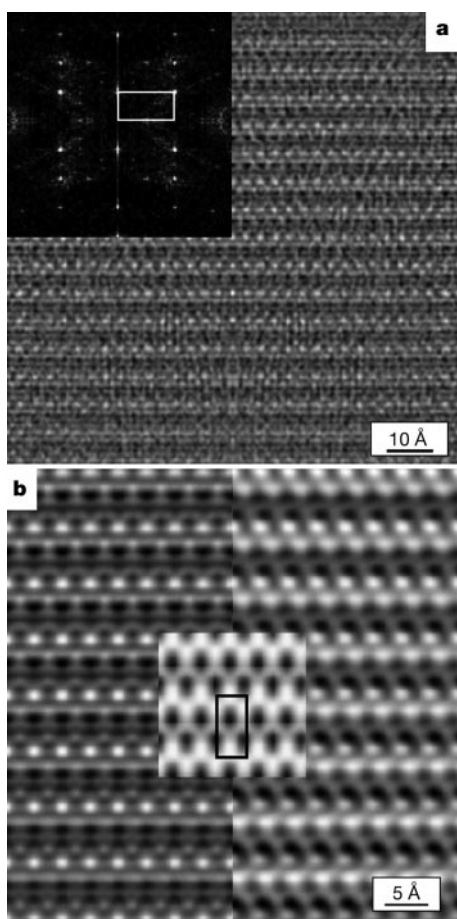


Figure 2 High-resolution image of the 2×1 SrTiO₃ (001) surface. Image was obtained with the JEOL-4010 TEM at Argonne National Laboratory. **a**, The image after being Wiener-filtered²³ and the bulk spots digitally removed. The inset shows the corresponding power spectrum, with unit cell marked. **b**, Enlarged translation-averaged image of the 2×1 structure (left) and after filtering to separate translation/inversion domains on the top and bottom surface (right). The inset shows a multislice simulation of the structure (defocus –55 nm) with the unit cell marked. The strong black features correlate with the positions of Ti atoms at the surface.

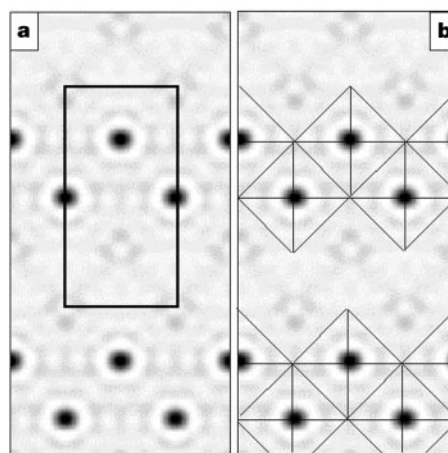


Figure 3 Theoretical direct methods solution of the (2×1) structure. **a**, Scattering potential map calculated from two-dimensional code for pm symmetry. **b**, Interpretation of the map in terms of TiO_x units. The primitive unit cell is indicated with solid lines ($a = 3.905$ Å, $b = 7.81$ Å, $\alpha = 90^\circ$). In both cases regions of high potential—that is, possible atomic sites—are black, and subsequent analysis demonstrated that these were Ti atoms.

a slightly distorted bulk-like TiO_2 -layer-terminated SrTiO_3 subsurface.

We note the retention of edge-shared TiO_5 units, yielding a charge-neutral reconstructed surface with the stoichiometry Ti_2O_4 . Preliminary results for the 4×2 and the 6×2 SrTiO_3 surface structures show similar structural motifs, and suggest similar surface rearrangements. These results indicate that edge-shared assembly of TiO_x units on the surface, analogous to creation of 'block structures' in $\text{Nb}_2\text{O}_{5-x}$ and other ReO_3 -type bulk oxide structures, provides additional modes for rearrangement to lower-energy configurations and controls the surface structure of SrTiO_3 .

We suggest that this structure formation rule can be extended to perovskites in general. We could compare this result for perovskite surfaces with the rules of sp^3 bonding for semiconductor surfaces, an idea first suggested for Si surfaces that quickly became widely used to explain surface structure formation in other semiconductors. This suggests that many of the concepts of bulk chemistry can also be applied to surfaces, coupled with minimizing surface dipole moments and maximizing coordination. Different specific results will occur, depending upon the bonding character and structural rules, for instance, delocalized electron states in metals, directional sp^2 or sp^3 in many semiconductors and block structures in oxides. □

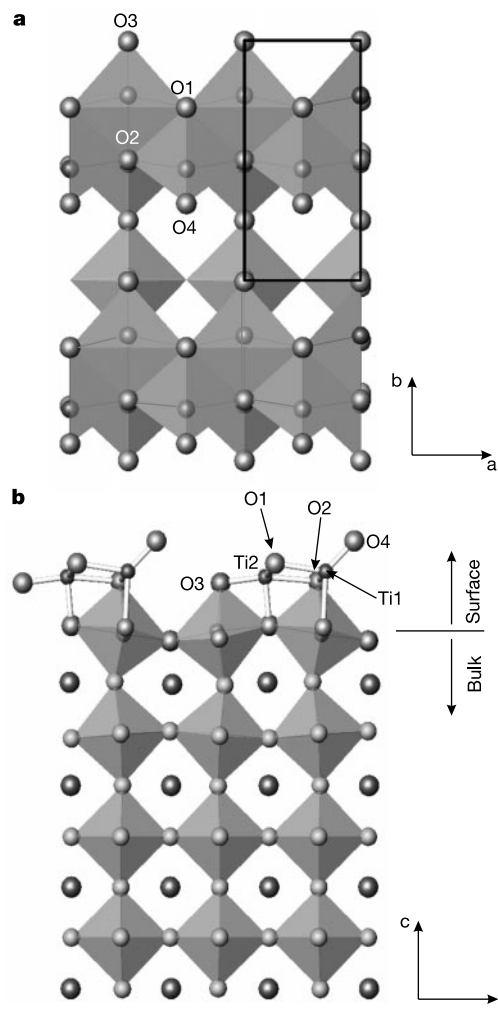


Figure 4 The 2×1 SrTiO_3 (001) surface structure. **a**, Top view; **b**, side view. TiO_x units are shown as partially filled octahedra on the surface; the primitive unit cell is indicated with solid lines ($a = 3.905 \text{ \AA}$, $b = 7.81 \text{ \AA}$).

Methods

Diffraction analysis

A series of negatives with exposure times varying from 0.5 s to 120 s were recorded for the 2×1 reconstruction using strategies that we have developed over the past decade²⁹. The negatives were digitized to eight bits with a $25\text{-}\mu\text{m}$ pixel size using an Optronics P-1000 microdensitometer, and intensities extracted using a cross-correlation technique³⁰, then averaged using a $p2mm$ Patterson plane group symmetry to yield 18 independent intensities. (Under the exposure conditions used, the intensity readout from the microdensitometer was proportional to the true intensities of the diffraction spots.) Because $p2mm$ Patterson symmetry may correspond to pm , pg , $p2mm$, $p2mg$ or $p2gg$ plane group symmetries, all were tested in the calculations. Final structure refinement was performed based on χ^2 , defined as:

$$\chi^2 = (N - M)^{-1} \sum [(I_{\text{meas}} - I_{\text{calc}})/\sigma]^2 \quad (1)$$

where I_{calc} is the calculated intensity, I_{meas} the measured intensity, N the number of data points, M the number of variable parameters, and σ the measurement error.

Electronic structure

DFT calculations were performed using the *ab initio* total-energy and molecular-dynamics program VASP (Vienna *ab initio* simulation package) developed at the Institut für Material-physik der Universität Wien^{24–27}, employing the generalized-gradient approximation²⁷. The VASP code uses ultrasoft pseudopotentials²⁶ and a plane-wave expansion of the electronic wave functions.

Received 22 May; accepted 23 July 2002; doi:10.1038/nature01010.

- Wolf, D. Reconstruction of NaCl surfaces from a dipolar solution to the Madelung problem. *Phys. Rev. Lett.* **68**, 3315–3318 (1992).
- Pashley, M. D. Electron counting model and its application to island structures on molecular-beam epitaxy grown GaAs(001) and ZnSe(001). *Phys. Rev. B* **40**, 10481–10487 (1989).
- Kawasaki, M. *et al.* Atomic control of the SrTiO_3 crystal-surface. *Science* **266**, 1540–1542 (1994).
- Cord, B. & Courths, R. Electronic study of SrTiO_3 surfaces by photoemission. *Surf. Sci.* **162**, 34–38 (1985).
- Jiang, Q. D. & Zegenhagen, J. $c(6 \times 2)$ and $c(4 \times 2)$ reconstruction of SrTiO_3 (001). *Surf. Sci.* **425**, 343–354 (1999).
- Naito, M. & Sato, H. Reflection high-energy electron diffraction study on the SrTiO_3 surface structure. *Physica C* **229**, 1–11 (1994).
- Jiang, Q. D. & Zegenhagen, J. SrTiO_3 (001)- $c(6 \times 2)$: A long-range, atomically ordered surface stable in oxygen and ambient air. *Surf. Sci.* **367**, L42–L46 (1996).
- Nishimura, T., Ikeda, A., Namba, H., Morishita, T. & Kido, S. Structure change of TiO_2 -terminated SrTiO_3 (001) surfaces by annealing in O_2 atmosphere and ultra-high vacuum. *Surf. Sci.* **421**, 273–278 (1999).
- Liang, Y. & Bonnell, D. A. Atomic structures of reduced SrTiO_3 (001) surfaces. *Surf. Sci. Lett.* **285**, L510–L516 (1993).
- Matsumoto, T., Tanaka, H., Kawai, T. & Kawai, S. STM-imaging of a SrTiO_3 (100) surface with atomic-scale resolution. *Surf. Sci.* **278**, L153–L158 (1992).
- Szot, K. & Speier, W. Surfaces of reduced and oxidized SrTiO_3 from atomic force microscopy. *Phys. Rev. B* **60**, 5909–5926 (1999).
- Padilla, J. & Vanderbilt, D. *Ab initio* study of SrTiO_3 . *Surf. Sci.* **418**, 64–70 (1998).
- Aruta, C. Structure of superconducting $[\text{BaCuO}_2]_2/[\text{CaCuO}_2]_n$ superlattices on SrTiO_3 (001) investigated by X-ray scattering. *Phys. Status Solidi A* **183**, 353–364 (2001).
- Droopard, R. *et al.* Development of high dielectric constant epitaxial oxides on silicon by molecular beam epitaxy. *Mater. Sci. Eng. B* **87**, 292–296 (2001).
- Marks, L. D. *et al.* Direct methods for surfaces. *Surf. Rev. Lett.* **5**, 1087–1106 (1998).
- Marks, L. D., Erdman, N. & Subramanian, A. Crystallographic direct methods for surfaces. *J. Phys. Condens. Matter* **13**, 10677–10688 (2001).
- Collazo-Davila, C., Grozea, D. & Marks, L. D. Determination and refinement of the $\text{Ag/Si}(111)-(3 \times 1)$ surface structure. *Phys. Rev. Lett.* **80**, 1678–1681 (1998).
- Marks, L. D., Sinkler, W. & Landree, E. A feasible set approach to the crystallographic phase problem. *Acta Cryst.* **55**, 601–612 (1999).
- Xu, P. & Marks, L. D. Intensities of surface diffraction spots in plan view. *Ultramicroscopy* **45**, 155–157 (1992).
- Collazo-Davila, C. *et al.* Design and initial performance of an ultrahigh vacuum Sample Preparation Evaluation Analysis and Reaction (SPEAR) system. *J. Microsc. Soc. Am.* **1**, 267–279 (1995).
- Marks, L. D. Wiener-filter enhancement of noisy HREM images. *Ultramicroscopy* **62**, 43–52 (1996).
- Marks, L. D. & Plass, R. Atomic-structure of $\text{Si}(111)-(5 \times 2)\text{-Au}$ from high resolution electron microscopy and heavy-atom holography. *Phys. Rev. Lett.* **75**, 2172–2175 (1995).
- Bengu, E. *et al.* Imaging the dimers in $\text{Si}(111)-(7 \times 7)$. *Phys. Rev. Lett.* **77**, 4226–4228 (1996).
- Kresse, G. & Hafner, J. *Ab-initio* molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
- Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
- Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue problem. *Phys. Rev. B* **41**, 7892–7895 (1990).
- Perdew, J. P. in *Electronic Structure of Solids '91* (eds Ziesche, P. & Eschrig, H.) 11 (Akademie, Berlin, 1991).
- Warschkow, O., Dyke, J. M. & Ellis, D. E. A divide-and-conquer implementation of the discrete variational DFT method for large molecular and solid systems. *J. Comp. Phys.* **143**, 70–89 (1998).
- Jayaram, G., Xu, P. & Marks, L. D. Atomic structure of the Si (001) 2×1 surface. *Phys. Rev. Lett.* **71**, 3489–3492 (1993).
- Xu, P., Jayaram, G. & Marks, L. D. Cross-correlation method for intensity measurement of transmission electron-diffraction patterns. *Ultramicroscopy* **53**, 15–18 (1994).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

HREM analysis was carried out at the Electron Microscopy Collaborative Research Center at Argonne National Laboratory. This work was supported by the EMSI program of the National Science Foundation and the US Department of Energy Office of Science at the Northwestern University Institute for Environmental Catalysis. M.A. acknowledges funding from the National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.D.M. (e-mail: l-marks@northwestern.edu).

Evidence from the AD 2000 Izu islands earthquake swarm that stressing rate governs seismicity

Shinji Toda*, Ross S. Stein† & Takeshi Sagiya‡

* Active Fault Research Center, Geological Survey of Japan, AIST, Tsukuba 305-8567, Japan

† US Geological Survey, MS 977, Menlo Park, California 94025, USA

‡ Geographical Survey Institute, Tsukuba 305-0811, Japan

Magma intrusions and eruptions commonly produce abrupt changes in seismicity far from magma conduits^{1–4} that cannot be associated with the diffusion of pore fluids or heat⁵. Such ‘swarm’ seismicity also migrates with time, and often exhibits a ‘dog-bone’-shaped distribution^{3,4,6–9}. The largest earthquakes in swarms produce aftershocks that obey an Omori-type (exponential) temporal decay^{10–12}, but the duration of the aftershock

sequences is drastically reduced, relative to normal earthquake activity^{7,13}. Here we use one of the most energetic swarms ever recorded to study the dependence of these properties on the stress imparted by a magma intrusion^{8,11,14,15}. A 1,000-fold increase in seismicity rate and a 1,000-fold decrease in aftershock duration occurred during the two-month-long dyke intrusion. We find that the seismicity rate is proportional to the calculated stressing rate, and that the duration of aftershock sequences is inversely proportional to the stressing rate. This behaviour is in accord with a laboratory-based rate/state constitutive law¹⁶, suggesting an explanation for the occurrence of earthquake swarms. Any sustained increase in stressing rate—whether due to an intrusion, extrusion or creep event—should produce such seismological behaviour.

The swarm struck the Izu volcanic islands 150 km south of Tokyo, producing 7,000 shocks with magnitude $M \geq 3$, and five $M \geq 6$ shocks; the total seismic energy release was 1.5×10^4 J, nearly an order of magnitude larger than the swarms that occurred in 1965–67 at Matsushiro, Japan, or in 1980 in Long Valley, California. The Izu swarm was accompanied by five phreatic eruptions of Miyakejima. Seismic activity began on 26 June with a shallow, dense swarm under Miyakejima, and migrated northwest (Fig. 1a). In July, the swarm developed northern and southern lobes (Fig. 1c, d). Earthquakes with $M = 6.4$ (1 July), $M = 6.1$ (9 July) and $M = 6.0$ (18 August) struck near the centre of the swarm, and $M = 6.3$ (15 July) and $M = 6.4$ (30 July) shocks struck ~ 25 km from the centre. Although it is generally assumed that the duration of aftershock sequences is proportional to mainshock magnitude¹⁰, aftershocks of these $M \approx 6$ strike-slip earthquakes persisted for as little as a day, whereas aftershocks of $M = 6$ events normally last several years in the Izu islands (Fig. 2a). The distance between Kozushima and Shikinejima gradually extended by 0.85 m until 23 August, when the seismicity and rapid displacement ceased¹⁷ (Fig. 3a).

The seismicity and deformation data are most compatible with a laterally propagating dyke intrusion, a process common to the Izu peninsula^{8,9,18}. We assume that a vertical dyke propagated to its full length in the first week (Fig. 1a), and then opened continuously for seven weeks. We infer 20 m of expansion over a depth extent of

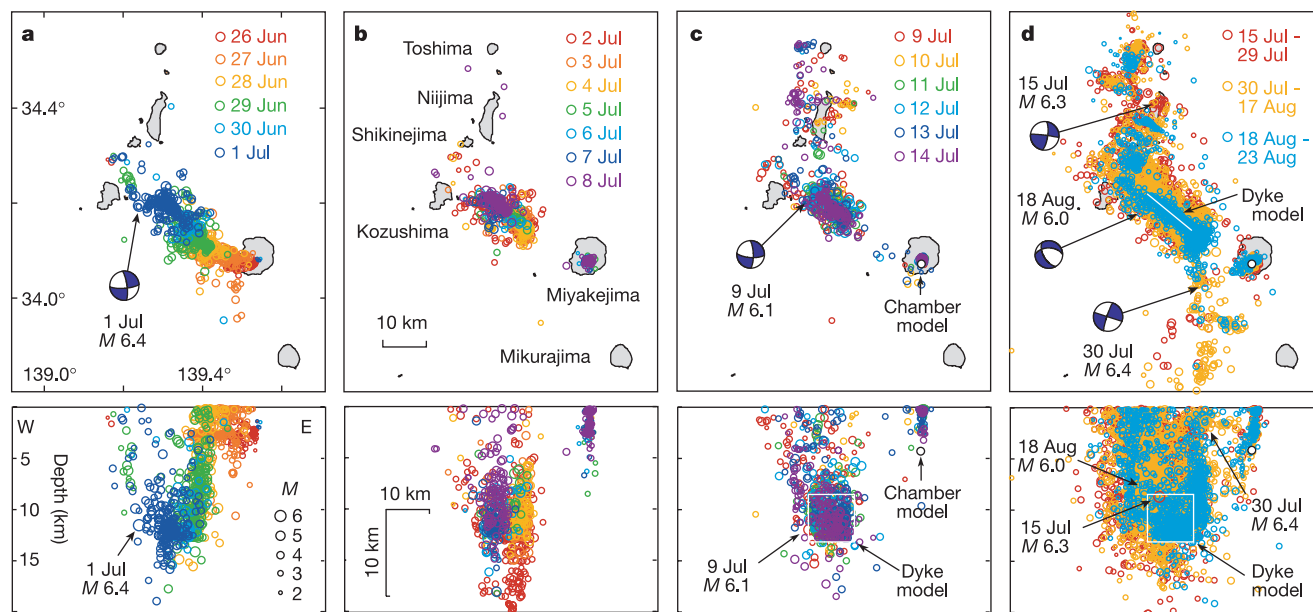


Figure 1 Swarm evolution in map view (top panels) and cross-section (bottom panels). Data from Earthquake Research Institute, University of Tokyo (ERI). Off-dyke seismicity

appears within 3 d (a), and expands substantially after two weeks (b–d). Geometry of the inferred dyke and magma chamber is shown in c, d.

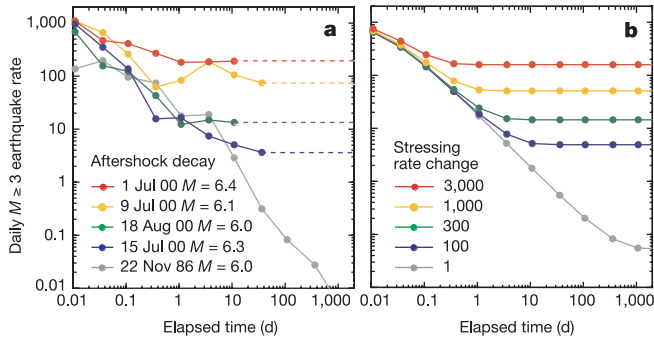


Figure 2 Observed and predicted aftershock decay. **a**, Observed decays for the four largest shocks during the Izu swarm, and for the nearest $M = 6$ shock that occurred well before the swarm (20 km east of Toshima, grey). The 1 July, 9 July and 18 August shocks struck near the dyke; the 15 July shock occurred 20 km north of the dyke. Aftershocks were counted in a 30-km-long cube centred on each hypocentre; curves are dashed after the swarm ended. **b**, Predicted aftershock decay as a function of the stressing rate change. The similarity between **a** and **b** suggests that aftershock decay is controlled principally by the stressing rate. Curves in **b** were generated from equations (12) and (14) in ref. 16, assuming a background stressing rate $\dot{\tau}_r = 0.1 \text{ bar yr}^{-1}$, $A\sigma_n = 0.1 \text{ bar}$, and a 1-bar mean stress gain imparted to aftershocks of each $M = 6$ shock. The modelled earthquake rate is given by the stressing rate change times the observed (1980–99) background $M \geq 3$ seismicity rate of 0.05 d^{-1} . The curves are insensitive to the stress change except for the first few hours, but they are sensitive to the background stressing and seismicity rates for longer periods. Stressing rate changes in **b** are chosen to correspond to those estimated for the observed mainshocks in **a**, based on their location, depth, and time of occurrence in the sequence.

8–13 km, and a volume increase of $\sim 1.5 \text{ km}^3$, corresponding to a geodetic moment of $\sim 5 \times 10^{19} \text{ N m}$ (Fig. 3b). GPS (Global Positioning System) vectors on Miyake also indicate a magma chamber deflation of 0.12 km^3 at 4 km beneath Miyake during the swarm¹⁷.

We use the dyke expansion model to test whether the swarm seismicity is controlled by stress transferred by the intrusion. Given the background stressing rate ($\sim 0.1 \text{ bar yr}^{-1}$) inferred from the strain rate¹⁹, and stressing rates of up to 150 bar yr^{-1} calculated during the intrusion (Fig. 4a), we infer a 1,500-fold gain in the shear

Table 1 Expected number of damaging Izu swarm earthquakes

Magnitude	Location	Background rate (yr^{-1})	Stressing rate change	Number of shocks	
				Expected*	Observed
$M \geq 6$	Total area	0.09	400	5.8	5
	Near dyke	0.005	3,250	2.6	3
$M \geq 5$	Total area	0.7	400	45	41
	Near dyke	0.04	3,250	21	33

There were seven shocks with $M \geq 6$ from 1926 to 1999 (0.09 yr^{-1}) in the area shown in Fig. 4. The calculated mean swarm stressing rate for this area is 40 bar yr^{-1} , and the background rate is $\sim 0.1 \text{ bar yr}^{-1}$, so the rate change is 400. The $20 \text{ km} \times 20 \text{ km}$ near-dyke region comprises the central 6% of the area, so the background rate of near-dyke $M \geq 6$ shocks is 0.005 yr^{-1} .

*Data in this column is given by column 3 entry $\times$ column 4 entry $\times$ the swarm period (0.16 yr).

stressing rate near the dyke. Seismicity is generally abundant where the stressing rate is calculated to have increased, and rare where it decreased. The calculated stressing rate change (Fig. 4a) is correlated with the observed seismicity rate change (Fig. 4b), suggesting a causal relation.

An alternative hypothesis is that ground water heated by the intrusion diffused outward along pre-existing fractures in the ‘dog-bone’ lobes, promoting earthquakes by raising pore pressure or temperature. The penetration of the diffusive front is a function of the ratio of the thermal to hydraulic diffusivities⁵. For appropriate values of the thermal ($10^{-6} \text{ m}^2 \text{ s}^{-1}$) and hydraulic ($10^{-2} \text{ m}^2 \text{ s}^{-1}$) diffusivities, the pore pressure increase is negligible beyond 250 m in the first 20 d (ref. 5). Even for highly fractured basalt with permeability as low as 10^{-15} m^2 (ref. 20) at the 5–15 km depth of seismicity, the pulse would reach no more than 2.5 km in 20 d, by which time several $M \approx 6$ shocks and numerous smaller ones had appeared 25 km from the Izu dyke (Fig. 1c, d). Because similar aftershock durations are seen for the events of 18 August (3 km from the dyke) and 15 July (20 km from the dyke), and the decay of the 18 August event is much longer than for the 1 July shock despite a similar proximity to the dyke (Fig. 2a), a temperature dependence of aftershock duration¹¹ is not evident. Thus, neither pore-fluid nor heat diffusion is likely to explain the remote triggering and shortened aftershock durations at Izu and elsewhere.

In contrast, the stressing-rate dependences of seismicity rate and aftershock duration are properties of a stress transfer model incorporating rate/state friction, for which the seismicity rate

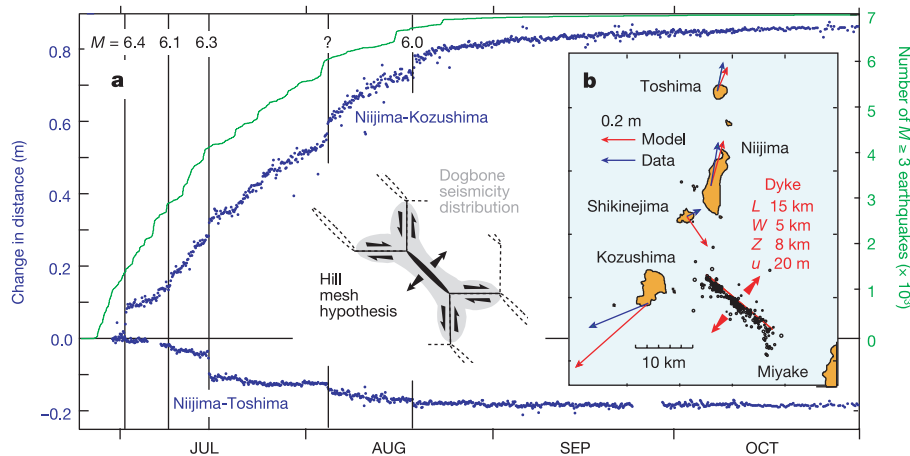


Figure 3 Intrusion geometry and deformation. **a**, GPS line-length changes¹⁷ and cumulative Japan Meteorological Agency (JMA) $M \geq 3$ shocks, with the observed ‘dog-bone’ pattern (grey) and Hill mesh explanation⁶ (black) inset. **b**, We use epicentres at 9–12 km depth located by ocean bottom seismometers to constrain the dyke position²²; vectors are net displacements; L , length; W , width; Z , upper depth; u , opening. The

volume lost to collapse of the summit caldera (0.60 km^3), magma chamber deflation (0.12 km^3), and eruptions (0.02 km^3)¹⁷ is about half that inferred for the dyke intrusion (1.5 km^3), suggesting that some magma migrated from greater depths. The inferred two-month intrusion rate is $\sim 300 \text{ m}^3 \text{ s}^{-1}$, 10 times that of the 1997 off-Izu intrusion⁹.

should be linearly related to the swarm stressing rate (Fig. 5). Although the data are noisy, the observed regression is quite close to the rate/state model (Fig. 4b inset). Aftershock duration t_a , the time until the rate of seismicity returns to what prevailed before the mainshock, is related to shear stressing rate $\dot{\tau}$ by $t_a = A\sigma_n/\dot{\tau}$, where $A\sigma_n$ is a constitutive parameter times the normal stress¹⁶, which we assume to be unchanged by the swarm¹⁵. The shortened aftershock duration is due to the increased level of background seismicity and to a change in the Omori decay exponent during the swarm (Fig. 2b). Relative to periods without intrusions, aftershock durations should decrease by a factor of $\sim 1,000$ above the dyke, and by a factor of ~ 100 in the off-dyke lobes, in rough accord with the data (Fig. 2a).

The 'dog-bone' pattern of seismicity seen at Izu, and to varying degrees elsewhere^{3,4,7-9,18}, can be explained by the transfer of shear stress from an expanding dyke to vertical faults in the surrounding crust (Fig. 4a). This pattern, as well as the prevalence of strike-slip focal mechanisms for all but mid-ocean-ridge swarms, were formerly explained by the Hill mesh⁶, an hypothesis that strike-slip faults link the dykes, accommodating and transferring the dyke expansion through the mesh (Fig. 3a inset). We instead suggest that seismicity occurs in these lobes whether or not connecting

faults exist, and that stress in each lobe can be relieved by right- and left-lateral faulting.

If swarm seismicity is indeed governed by the stressing rate, then the onset and rate of damaging earthquakes in a swarm can be forecast. The expected earthquake rate is the product of the stressing rate change, the background rate for shocks of a given magnitude, and the swarm duration. Retrospective agreement is good for both $M \geq 5$ and $M \geq 6$ Izu shocks (Table 1). Because seismicity reaches equilibrium more slowly at lower stressing rates (Fig. 5b), earthquakes appear to migrate away from the dyke. The near-dyke shocks should start within days of the intrusion (the first $M \geq 6$ shock struck 5 d into the swarm), whereas those off the dyke should be delayed by several weeks ($M \geq 6$ shocks began there after 20 d). The observed rate of seismicity migration is thus more consistent with stress transfer than with pore-fluid diffusion.

Whereas several previous studies have assumed that rate/state friction controls earthquake occurrence^{15,21}, the voluminous burst of earthquakes at Izu provides the strongest observational test yet of this constitutive law for the genesis of seismicity. Rate/state stress transfer furnishes a comprehensive explanation for distributed swarm seismicity, triggering and clustering, and offers the prospect that near-real-time analysis of seismic and GPS data may permit

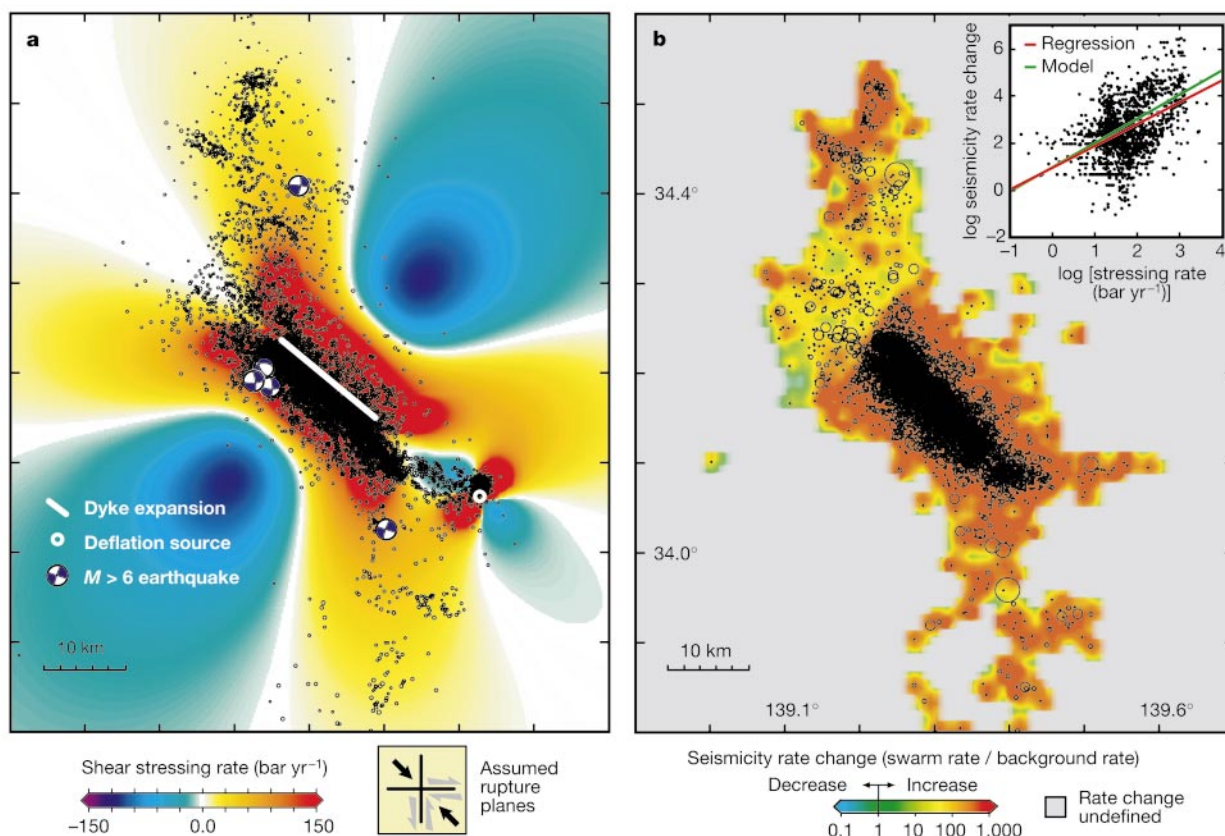


Figure 4 Calculated shear stressing rate and observed seismicity rate. **a**, Shear stressing rate caused by dyke expansion and Miyake deflation (modelled using Coulomb 2.2 (ref. 21) at depths of 2–8 km resolved on vertical planes, consistent with focal mechanisms). Seismicity (ERI, 26 June to 23 August 2000) is more common where the stressing rate is calculated to have increased. **b**, Seismicity rate change²¹ smoothed with a 1-km gaussian smoothing radius. The JMA catalogue is used because it is complete to $M \geq 3$ from 1985; 'swarm' is 59 d after 29 June 2000; 'background' is 15 yr before 29 June 2000. The average background rate was substituted for sites with at least one swarm shock but no background seismicity. The brevity of the swarm all but precludes observation of seismicity rate decreases. The inset shows the spatial regression of **a** on **b**. The model

(green) is found by combining equations (10) and (11) from ref. 16, for which $R/r = \dot{\tau}/\dot{\tau}_r$ once γ , the state variable for seismicity, reaches steady state; R and $\dot{\tau}$ are the swarm seismicity and stressing rates; r and $\dot{\tau}_r$ are the background seismicity and stressing rates. The model predicts $y = 1.0 \log x + 1.0$; the linear regression (red) yields $y = 0.93 \log x + 0.95$ with correlation coefficient 0.48, significant at the 99.99% confidence level for the effective number of observations (542). Synthetic seismicity with the property that the seismicity rate decreases inversely with distance from the dyke exhibits a much poorer fit to the rate/state model ($y = 0.34 \log x + 2.6$), with slope one-third of the observations or model. The correlation coefficient for the synthetic data regression is < 0.75 , but depends principally on the amount of introduced noise.

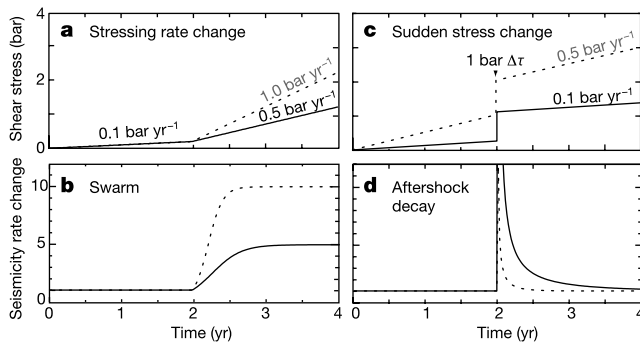


Figure 5 The rate/state effect of stress on seismicity. Details of the calculations are given in Methods. A change in the stressing rate (**a**) causes a swarm (**b**). A sudden stress change, $\Delta\tau$ (**c**), causes an aftershock sequence that decays inversely with time (**d**). The Izu swarm has several aftershock sequences embedded in it. Comparison of dashed and solid curves shows that the higher the stressing rate, the more quickly the seismicity rate reaches equilibrium. As the stressing rate change is highest close to the source, swarm seismicity appears to migrate away from an intrusion or creep site.

forecasts of damaging earthquakes during future swarms. Equally important, the Izu test suggests that a central unresolved problem of earthquake interaction—the response of seismicity to a large shock followed by viscoelastic rebound—is essentially a sudden stress change succeeded by a transient stressing rate change, which can be simulated by combining the two processes shown in Fig. 5. □

Methods

To calculate the updated daily seismicity rate, R , due to a stressing rate change (Fig. 5b), we seek the updated state variable γ from equation (11) of ref. 16, $R = r/(\gamma\dot{\tau})$, where $\dot{\tau}$ is the background stressing rate, and the background seismicity rate r is set to 1. At $t = 0$, γ is steady state, where $\gamma_{ss} = 1/(\dot{\tau})$. To evolve γ , we use equation (B17) of ref. 16, $\gamma = [\gamma_0 - \frac{1}{\dot{\tau}}] \exp\left[\frac{\dot{\tau}t}{A\sigma_n}\right] + \frac{1}{\dot{\tau}}$, where γ_0 is the state variable before each time step, and $\dot{\tau}$ is the stressing rate. For the response to a sudden stress change $\Delta\tau$ (Fig. 5d), $\gamma = \gamma_0 \exp\left(\frac{\Delta\tau}{A\sigma_n}\right)$, modified from equation (B11) in ref. 16. For the Izu swarm, we infer $A\sigma_n$ using the relation $A\sigma_n = t_a \dot{\tau}$. From Fig. 2a, t_a for the $M \approx 6$ shocks close to the dyke is ~ 0.3 d where the calculated stressing rate $\dot{\tau} \approx 150 \text{ bar yr}^{-1}$. The observed t_a for the background $M \approx 6$ shock in Fig. 2a is ~ 1 yr, and the background $\dot{\tau} \approx 0.1 \text{ bar yr}^{-1}$. Both estimates yield $A\sigma_n \approx 0.1$ bar, which we use here. The mean stressing rate in Fig. 4a is 32 bar yr^{-1} , and the mean t_a for the $M \approx 6$ shocks is ~ 3 d, for $A\sigma_n \approx 0.3$ bar, similar to a previous estimate²¹.

Received 16 April; accepted 19 July 2002; doi:10.1038/nature00997.

- Weaver, C. S., Grant, W. C., Malone, S. D. & Endo, E. T. The 1980 eruptions of Mt. St. Helens, Washington. *US Geol. Surv. Prof. Pap.* **1250**, 109–121 (1981).
- Mori, J., Eberhart-Phillips, D. & Harlow, D. H. *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines* (eds Newhall, C. G. & Punongbayan, R. S.) 371–382 (Univ. Washington Press, Seattle, 1996).
- Björnsson, A., Saemundsson, K., Einarsson, P. & Tryggvason, E. Current rifting episode in north Iceland. *Nature* **266**, 318–323 (1977).
- Dvorak, J. J. *et al.* Mechanical response of the south flank of Kilauea Volcano, Hawaii, to intrusive events along the rift systems. *Tectonophysics* **124**, 193–209 (1986).
- Delaney, P. T. Rapid intrusion of magma into wet rock: Groundwater flow due to pore pressure increases. *J. Geophys. Res.* **87**, 7739–7756 (1982).
- Hill, D. A model for earthquake swarms. *J. Geophys. Res.* **82**, 1347–1352 (1977).
- Klein, F. W., Einarsson, P. & Wyss, M. The Reykjanes peninsula, Iceland, earthquake swarm of September 1972 and its tectonic significance. *J. Geophys. Res.* **82**, 865–888 (1977).
- Ukawa, M. & Tsukahara, H. Earthquake swarms and dike intrusions off the east coast of Izu peninsula, central Japan. *Tectonophysics* **253**, 285–303 (1996).
- Aoki, Y., Segall, P., Kato, T., Cervelli, P. & Shimada, S. Imaging magma transport during the 1997 seismic swarm off the Izu peninsula, Japan. *Science* **286**, 927–930 (1999).
- Watanabe, K. On the duration time of aftershock activity. *Bull. Disast. Prev. Res. Inst. Kyoto Univ.* **39**, 1–21 (1989).
- Kisslinger, C. & Jones, L. Properties of aftershock sequences in southern California. *J. Geophys. Res.* **96**, 11947–11958 (1991).
- Gross, S. J. & Kisslinger, C. Tests of models of the aftershock rate decay. *Bull. Seismol. Soc. Am.* **84**, 1571–1579 (1994).
- Walter, A. W. & Weaver, C. S. Seismicity of the Coso Range, California. *J. Geophys. Res.* **85**, 2441–2458 (1980).
- Chouet, B. A. Long-period volcano seismicity: its source and use in eruption forecasting. *Nature* **380**, 309–316 (1996).
- Dieterich, J., Cayol, V. & Okubo, P. The use of earthquake rate changes as a stress meter at Kilauea volcano. *Nature* **408**, 457–460 (2000).
- Dieterich, J. A constitutive law for rate of earthquake production and its application to earthquake clustering. *J. Geophys. Res.* **99**, 2601–2618 (1994).

- Nishimura, T. *et al.* Crustal deformation caused by magma migration in the northern Izu Islands, Japan. *Geophys. Res. Lett.* **28**, 3745–3748 (2001).
- Okada, Y. & Yamamoto, E. A model for the 1989 seismo-volcanic activity off Ito, central Japan, derived from crustal movement data. *J. Phys. Earth* **39**, 177–195 (1991).
- Sagiya, T., Miyazaki, S. & Tada, T. Continuous GPS array and present-day crustal deformation of Japan. *Pure Appl. Geophys.* **157**, 2303–2322 (2000).
- Ingebritsen, S. E. & Sanford, W. E. *Groundwater in Geologic Processes* 16 (Cambridge Univ. Press, Cambridge, 1998).
- Toda, S., Stein, R. S., Reasenberg, P. A. & Dieterich, J. H. Stress transferred by the $M_w = 6.5$ Kobe, Japan, shock: Effect on aftershocks and future earthquake probabilities. *J. Geophys. Res.* **103**, 24543–24565 (1998).
- Sakai, S. *et al.* Magma migration from the point of view of seismic activity in the volcanism of Miyakejima Island in 2000. *J. Geogr.* **110**, 145–155 (2001).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank J. Dieterich for guidance and insight; V. Cayol, B. Chouet, D. Hill and W. Thatcher for comments and suggestions; and M. Kikuchi, Y. Yamanaka, S. Sakai and H. Tsuruoka for help in the initial stages of this study. We used ERI, JMA and GSI data, and seismicity-rate computer codes by P. Reasenberg.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.T. (e-mail: s-toda@aist.go.jp).

Pretender punishment induced by chemical signalling in a queenless ant

Thibaud Monnin[†], Francis L. W. Ratnieks^{*}, Graeme R. Jones[‡] & Richard Beard[‡]

^{*} Laboratory of Apiculture and Social Insects, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

[‡] Chemical Ecology Group, School of Chemistry and Physics, University of Keele, Keele, Staffordshire ST5 5BG, UK

Animal societies are stages for both conflict and cooperation. Reproduction is often monopolized by one or a few individuals who behave aggressively to prevent subordinates from reproducing (for example, naked mole-rats¹, wasps² and ants³). Here we report an unusual mechanism by which the dominant individual maintains reproductive control. In the queenless ant *Dinoponera quadricaps*, only the alpha female reproduces. If the alpha is challenged by another female she chemically marks the pretender who is then punished⁴ by low-ranking females. This cooperation between alpha and low-rankers allows the alpha to inflict punishment indirectly, thereby maintaining her reproductive primacy without having to fight.

Queenless ponerine ants have evolutionarily lost the morphological queen caste⁵. All females are workers who can potentially mate and reproduce sexually (mated workers are called gamergates)⁶. Colonies of *D. quadricaps* have, on average, 80 adult workers and a single gamergate⁶, who has the alpha rank in a near-linear dominance hierarchy of about 3–5 high-ranking workers⁷. High-rankers are hopeful reproductives. They do little work, and one of them, usually the beta, replaces the gamergate if she dies⁷. Workers with lower ranks work, and are little involved in dominance

[†] Present address: Laboratoire d'écologie CNRS UMR 7625, Université Pierre et Marie Curie, 7 quai Saint Bernard, 75 005 Paris, France.

interactions. Males, as in other social Hymenoptera⁸, play no active role in colony life. Mating experiments show that the gamergate is mated to a single unrelated male⁶, so that workers are either daughters of the gamergate, or—if gamergate replacement has recently occurred—daughters and sisters of the new gamergate.

A high-ranker can enhance her inclusive fitness by overthrowing the gamergate, rather than by waiting for her to die naturally, because the high-ranker is more related to her own offspring than to those of another high-ranker (Methods, Benefits of attempted gamergate replacement to pretender). When a high-ranker challenges the gamergate, the two ants may engage in short fights and chase each other inside the nest, jostling others and trampling brood⁷. Bouts of chasing are interspersed with periods of relative calm, during which the gamergate may rub her sting against the pretender⁷ ('sting smearing', Fig. 1a). Following sting smearing, the pretender is often immobilized by low-ranking workers (ref. 7, and C. Peeters, personal communication) (Fig. 1b). Immobilization can last several days, and typically results in the pretender losing her high rank. It is not clear why punishment causes loss of rank, but it is probably a combination of the stress caused by immobilization and being prevented from performing dominance behaviours. Occasionally the immobilized individual is killed outright. Here we experimentally test the hypothesis that the gamergate can trigger immobilization of a pretender by a distinct chemical signal.

When performing sting smearing, the gamergate marks the pretender with chemicals from the Dufour's gland, which empties via the sting⁹. We smeared beta workers with the Dufour's gland

content of either a gamergate, or another beta or a low-ranker, and then recorded the duration of immobilization induced (Methods, Immobilization trial). Immobilization cannot be confused with other aggressive behaviours (Fig. 1b). We applied extracts to betas because beta is the rank most often naturally smeared by the gamergate⁷.

Gamergate glands induced immobilization significantly more often and for longer than beta or low-ranker glands (Fig. 2), and with more damaging consequences. Betas lost high rank significantly more often following smearing with gamergate gland than with gland of betas and low-rankers (4 of 9 trials versus 2 of 27 trials; $P = 0.0245$, Fisher's exact test, one-tailed). Because gland extracts were all from non-nestmates (Methods, Immobilization trial), our results also show that the signal is neither colony specific nor unique to each gamergate.

Chemical analyses (by gas chromatography/mass spectrometry, GC/MS) show that Dufour's glands contain mostly hydrocarbons (alkanes, alkenes and methyl-branched alkanes with 15–30 carbon atoms), and that gamergate glands differ from beta and low-ranker glands. Gamergate glands ($n = 12$) have significantly more hydrocarbons than low-rankers' ($n = 9$, summed area of the 30 peaks of highest abundance is 3.2 times higher in gamergates, Mann–Whitney U -test: $z = -2.487$, $P = 0.013$) but not more than beta glands ($n = 7$, summed area is 1.9 times higher in gamergates but $z = -1.268$, $P = 0.205$). Furthermore, gamergate glands contain a distinctive chemical mixture with a higher proportion of high-molecular-mass hydrocarbons than low-ranker glands, with beta glands intermediate (Figs 3 and 4). This is in agreement with the results of the bioassay, and suggests that Dufour's gland chemicals are a pheromonal signal that induces immobilization, and that only the gamergate produces sufficient quantities or the correct composition or both.

The use of a chemical signal to trigger immobilization, and hence to control reproduction, raises the question of why other females do not produce and apply this signal. One possibility is that the signal can only be produced by fully fertile individuals^{10,11}, whereas betas have only partially active ovaries at most¹². Alternatively, counter-

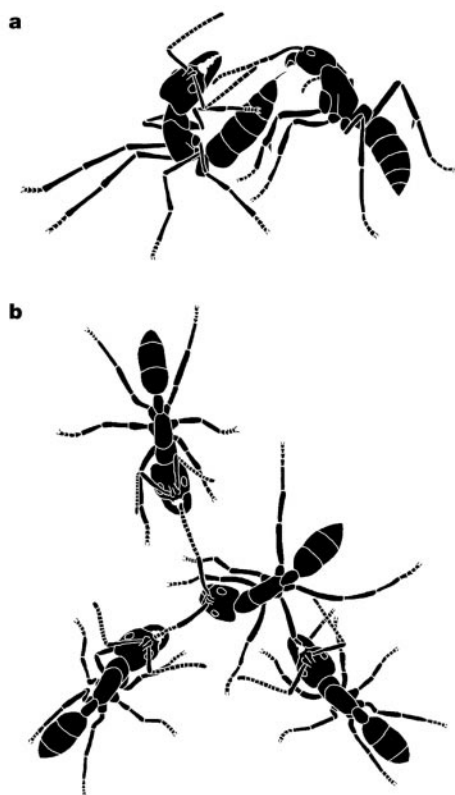


Figure 1 Dominance interactions. **a**, Sting smearing. The gamergate (left) approaches the pretender, usually from behind or from the side, briefly rubs her sting against the pretender without inserting the sting, and then runs away. Sting smearing differs from actual stinging. A stinging worker, typically a forager, bites and holds the prey and then inserts the sting into the prey to inject venom. **b**, Immobilization. One to six low-ranking workers bite and hold the appendages of the pretender for up to 3–4 days with workers taking turns (modified from ref. 7).

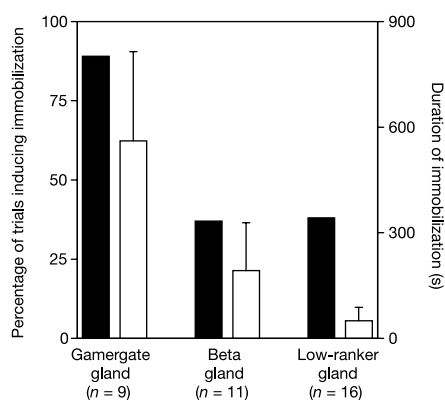


Figure 2 Bioassay results. Percentage of trials where betas were immobilized (black columns, left-hand y axis) following topical application of Dufour's gland content from either a gamergate, beta, or low-ranking worker, and average duration of immobilization (mean $\pm$ s.e.m., white columns, right-hand y axis). Gamergate gland triggers immobilization more often (χ^2 for the three groups, 7.263; degrees of freedom d.f. = 2; $P < 0.05$. Fisher exact test, one-tailed: gamergate versus beta, $P = 0.025$; gamergate versus low-ranker, $P = 0.017$) and induces longer immobilization (Kruskal–Wallis ANOVA for the three groups: $H = 9.151$, d.f. = 2, $P = 0.0103$; Mann–Whitney U -test: gamergate versus beta: $z = -2.198$, $P = 0.028$; gamergate versus low-ranker: $z = -2.899$, $P = 0.0037$). Beta and low-ranker glands do not differ in either effect (Fisher exact test, $P > 0.05$; Mann–Whitney U -test: $z = -0.285$, $P > 0.05$).

feiting the signal could be evolutionarily unstable, because it would result in repeated gamergate replacement. This would be costly to low-rankers, because it would reduce their genetic relatedness to the gamergate's offspring. Low-rankers would, therefore, benefit by ignoring the signal. Furthermore, frequent gamergate replacement would presumably decrease colony productivity.

Our results are, to our knowledge, the first to implicate the Dufour's gland in the resolution of intracolony reproductive conflict in ants. There are, however, several precedents for the importance of the Dufour's gland in ant aggression^{9,13}. Slave-makers and workerless social parasites trigger panic and aggression among host workers with Dufour's gland chemicals^{14,15}. Queen *Leptothorax gredleri* usurp conspecific nests by smearing resident queens with Dufour's gland secretions that induce resident workers to attack their own queens¹⁶. These propaganda substances manipulate worker behaviour in a non-adaptive way, as responding workers lose inclusive fitness¹⁴⁻¹⁶. This contrasts sharply with *D. quadricaps*, where workers who prevent gamergate replacement in response to the gamergate signal increase their inclusive fitness.

Immobilization also occurs in the queenless ponerine ants *Gnamptogenys menadensis*¹⁷ and *Harpegnathos saltator*¹⁸ (reviewed in ref. 19), but is not triggered by the gamergate. Both species have several gamergates per colony, and recruitment of additional gamergates does not result in the overthrow of current gamergates. Low-rankers probably regulate gamergate number to maintain an efficient balance between egg-layers and workers, given that gamergates work less. Immobilization is directed against workers activat-

ing their ovaries^{17,18}, who are probably detected by unavoidable chemical cues of ovarian activity²⁰⁻²², rather than by an externally applied signal as in *D. quadricaps*.

Our results show sophisticated cooperation between the gamergate and low-rankers in preventing gamergate overthrow by a pretender. Cooperation is favoured because both the gamergate and low-rankers are more related to the current gamergate's offspring than to the pretender's potential offspring (Methods, Benefits of cooperative punishment to gamergate and low rankers). The cooperation integrates the ability of the gamergate to signal the identity of the pretender with the collective physical power of low-rankers to immobilize her. Immobilization in *D. quadricaps* precisely fits the definition of punishment: "individuals (or groups) commonly responding to actions likely to lower their fitness with behaviour that reduces the fitness of the instigator and discourages or prevents him or her from repeating the initial action"⁴. Loss of high rank is costly, as the punished pretender loses any hope of future reproduction (only high-rankers can replace the gamergate⁷). Immobilization in *D. quadricaps* is perhaps the clearest yet demonstration of punishment in animal societies. Similar forms of punishment, where low-rankers prevent replacement of the reproductive, could potentially occur in many insect and vertebrate societies with totipotent individuals (for example, Polistinae and Stenogastrinae wasps, Halictidae bees, cooperatively breeding birds and mammals^{23,24}). In addition, because immobilization is a mechanism by which workers prevent other workers, the pretenders, from reproducing, it is also a novel form of worker policing^{7,19,25}, analogous to

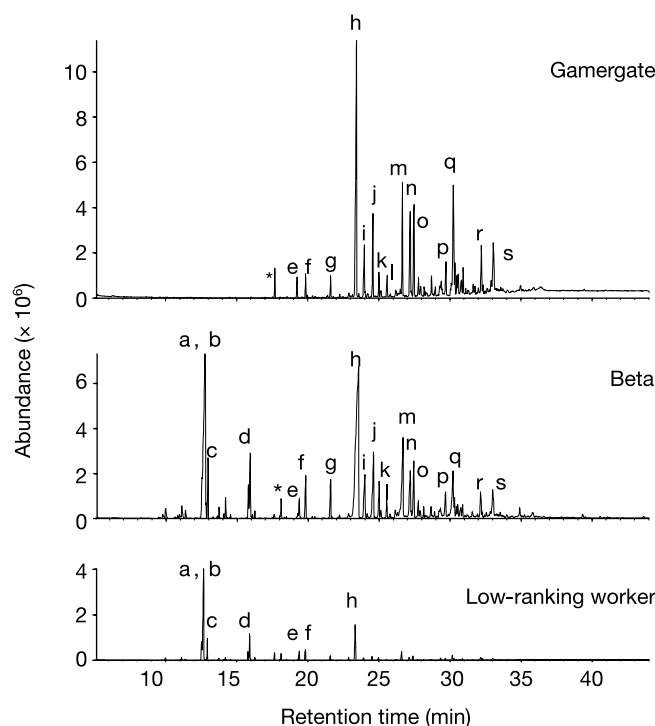


Figure 3 Chromatograms showing differences in Dufour's gland contents of workers of different ranks. Many of the hydrocarbons abundant in gamergates are rare in low-rankers and vice versa. Betas have intermediate profiles ($n = 12$ gamergates, 7 betas and 9 low-rankers). The major compounds are: heptadecadiene (a); (Z)8-heptadecene (b); *n*-heptadecane (c); (Z)9-nonadecene (d); (Z)9-heneicosene (e); *n*-heneicosane (f); *n*-docosane (g); *n*-tricosane (h); 7-,9-,11-methyltricosane (i); 3-methyltricosane (j); *n*-tetracosane (k); 11-methyltetracosane (l); *n*-pentacosane (m); 9-,11-methylpentacosane (n); 5-methylpentacosane (o); *n*-heptacosane (p); 9-,11-,13-methylheptacosane (q); nonacosane (r); and 4-,5-,6-octacosanone (s). Chemicals a-d are characteristic of betas

and low-rankers, e-h are common to all groups, and i-s are characteristic of gamergates and betas. Chemicals present in lower quantities include: *n*-pentadecane; 9-hexadecene; *n*-hexadecane; octadecene; nonadecadiene; *n*-nonadecane; octadecenal; tricosene; pentacosene; 3-methylpentacosane; hexacosene; *n*-hexacosane; 11-, 13-methylhexacosane; 5-methylhexacosane; heptacosene; 7-methylheptacosane; 5-methylheptacosane; 3-methylheptacosane; octacosene; 5-, 9-, 5-, 15-, 5-, 17-dimethylheptacosane; *n*-octacosane; 9-,11-,13-methyloctacosane; *n*-nonacosane; 9-,11-,13-,15-methylnonacosane; *n*-triacontane; and 9-,11-methyltriacontane. Asterisk indicates the contaminant isophthalate.

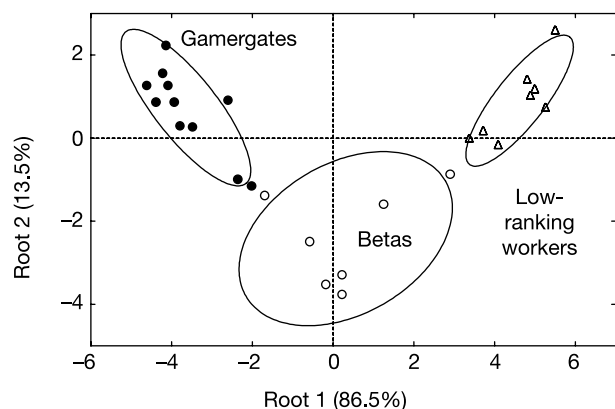


Figure 4 Discriminant analysis of Dufour's glands. The analysis separates gamergates, betas and low-rankers ($n = 12, 7$ and 9 , respectively), with no overlap of 95% confidence ellipses. The analysis was performed on the 5 factors of highest value resulting from a principal component analysis of the standardized proportions²⁰ of the 30 hydrocarbon peaks representing 0.5% or more of the total area. These 5 factors represent 92.9% of the total variance of the 30 hydrocarbon peaks.

the killing of worker-laid eggs in honeybees^{26,27} and wasps^{28,29} but acting in the context of policing breeder replacement rather than male production. Another difference is that policing by immobilization directly targets the 'selfish' worker who tries to reproduce, whereas policing by egg-killing targets only the products of selfish reproduction. □

Methods

Study animals

18 colonies of *D. quadricaps* were collected at the State University of Feira de Santana (Bahia, Brazil) in February 1998, February 2000 and October 2000. All individuals were individually marked (in Sheffield), and dominance interactions were recorded to determine social ranks. 'Blocking' and 'gaster rubbing' behaviours readily allow alpha and beta ants to be identified⁷.

Benefits of attempted gamergate replacement to pretender

High-rankers are usually daughters of the gamergate. They are not directly in conflict over reproduction with a mother gamergate, because they are as related to her offspring (0.25 brothers, 0.75 sisters), on average, as to their own offspring (0.5). However, a high-ranker can benefit by early replacement of a mother gamergate as a result of competition with other high-rankers, because she is more related to her own offspring (0.5) than to a sister's offspring (0.375)⁷. Additionally, a high-ranker would benefit from gamergate overthrow if, following recent gamergate replacement or colony fission, the gamergate is a sister (own offspring 0.5 versus sister's offspring 0.375).

Benefits of cooperative punishment to gamergate and low-rankers

The gamergate and low-rankers cooperate to prevent gamergate replacement because both parties are more related to the current gamergate's offspring than to the pretender's potential offspring (gamergate: own offspring 0.5 versus daughter's 0.25; low-rankers: mother gamergate's 0.5 versus sister's 0.375)⁷. Cooperation is still favoured if gamergate replacement has recently occurred. The new gamergate is then the low-rankers' sister, and pretenders are either sisters or nieces to low-rankers. Low-rankers favour the gamergate because they are equally or more related to the gamergate's offspring than to the pretenders'. Low-rankers additionally benefit from preventing replacement, because replacement itself would be costly. In particular, there would be a delay in brood rearing, because it takes approximately 6 weeks for a replacement alpha to mate⁶ and activate her ovaries fully³⁰. Furthermore, the colony would lose one worker, approximately 1% of the workforce, as the gamergate dies when overthrown⁷.

Immobilization trial

Dufour's glands of gamergates, betas and low-rankers were dissected ($n = 9, 11$ and 16 , respectively). Gamergate glands were thick, beige-brown and filled with a yellow oily secretion whereas beta and low-ranker glands were thin and pale-coloured. Histological studies show that gamergate glands have a thicker epithelium and biosynthetically more active cells than worker glands (J. Billen, personal communication; T.M. and F.L.W.R., unpublished work). The Dufour's glands, tubes approximately $10 \text{ mm} \times 1 \text{ mm}$ in gamergates, were cut in half across the length. One half was used for chemical analysis (below). The other half was applied on the cuticle of a beta worker from another colony to simulate sting smearing. We could not smear a beta with nestmate Dufour's glands because

it is impossible to treat a beta with her own gland. In addition, using the gland of the colony's gamergate would orphan the colony and increase worker aggressiveness⁷. Before application of the gland, the beta was chilled for 3–5 minutes to prevent her from struggling. The beta was then returned to her nest, where she rapidly warmed up and was video-recorded for 30 min. When several workers simultaneously immobilized the beta, we summed the durations of immobilization inflicted by each worker, to reflect the higher intensity of immobilization. Betas were used in 1–8 trials carried out at 1–121-day intervals. The beta was reused soon afterwards only when the previous trial had resulted in zero or little immobilization. The venom gland also empties by the sting. The destructiveness of the bioassay and the limited supply of gamergates prevented us from studying the effect of venom. However, venom is unlikely to be involved in sting smearing and gamergate behaviour. The venom sac is full in foragers, who use venom to kill large arthropod prey, but empty in gamergates (T.M., unpublished work).

Chemical analysis

Half of each Dufour's gland was extracted in 0.5 ml hexane. 100 μl of the extract was dried by evaporation under nitrogen flow, re-dissolved in 2 μl hexane and analysed with an high performance (HP) 5890 GC/MS apparatus fitted with a $30 \text{ m} \times 0.25 \text{ mm}$ internal diameter column covered with a 5% diphenyl-95% polysiloxane phase (1.0 μm thickness, Rtx-5, Restek). The carrier gas was helium (1 ml min^{-1}), and injections were splitless. The temperature was initially kept at 100 °C for 3 min, then increased by 15 °C min^{-1} to 170 °C, then by 5 °C min^{-1} to 300 °C and finally kept at 300 °C for 10 min. Linear and methyl-branched compounds were identified by retention times, equivalent chain length, NIST MS library and common fragmentation patterns. These identifications were not confirmed by co-injection of reference compounds and are, therefore, tentative. Double bounds were determined by derivatization with dimethyl disulphide, and their stereochemistry was determined by equivalent chain length and co-injection. Chirality of methyl-branched compounds was not determined. The numbers of individuals used for the bioassay and for chemical analysis vary, because some individuals were available for GC/MS when no colony was available for the bioassay (for example, a beta had been replaced recently, or the gamergate had been dissected).

Received 27 March; accepted 11 June 2002; doi:10.1038/nature00932.

- Clarke, F. M. & Faulkes, C. G. Dominance and queen succession in captive colonies of the eusocial naked mole-rat, *Heterocephalus glaber*. *Proc. R. Soc. Lond. B* **264**, 993–1000 (1997).
- Röseler, P.-F. In *The Social Biology of Wasps* (eds Ross, K. G. & Matthews, R. W.) 309–335 (Cornell Univ. Press, Ithaca, 1991).
- Heinze, J., Hölldobler, B. & Peeters, C. Conflict and cooperation in ant societies. *Naturwissenschaften* **81**, 489–497 (1994).
- Clutton-Brock, T. H. & Parker, G. A. Punishment in animal societies. *Nature* **373**, 209–216 (1995).
- Peeters, C. in *Queen Number and Sociality in Insects* (ed. Keller, L.) 234–261 (Oxford Univ. Press, Oxford, 1993).
- Monnin, T. & Peeters, C. Monogyny and regulation of worker mating in the queenless ant *Dinoponera quadricaps*. *Anim. Behav.* **55**, 298–306 (1998).
- Monnin, T. & Peeters, C. Dominance hierarchy and reproductive conflicts among subordinates in a monogynous queenless ant. *Behav. Ecol.* **10**, 323–332 (1999).
- Wilson, E. O. *The Insect Societies* (Harvard Univ. Press, Cambridge, Massachusetts, 1971).
- Billen, J. & Morgan, E. D. in *Pheromone Communication in Social Insects. Ants, Wasps, Bees, and Termites* (eds Vander Meer, R. K., Breed, M. D., Espelie, K. E. & Winston, M. L.) 3–33 (Westview, Boulder, 1998).
- Keller, L. & Nonacs, P. The role of queen pheromones in social insects: queen control or queen signal? *Anim. Behav.* **45**, 787–794 (1993).
- Ortius, D. & Heinze, J. Fertility signalling in queens of a North American ant. *Behav. Ecol. Sociobiol.* **45**, 151–159 (1999).
- Monnin, T. & Peeters, C. Cannibalism of subordinates' eggs in the monogynous queenless ant *Dinoponera quadricaps*. *Naturwissenschaften* **84**, 499–502 (1997).
- Hölldobler, B. & Wilson, E. O. *The Ants* (Harvard Univ. Press, Cambridge, Massachusetts, 1990).
- Regnier, F. E. & Wilson, E. O. Chemical communication and "propaganda" in slave-maker ants. *Science* **172**, 276–269 (1971).
- Allies, A. B., Bourke, A. F. G. & Franks, N. R. Propaganda substances in the cuckoo ant *Leptothorax kutteri* and the slave-maker *Harpagoxenus sublaevis*. *J. Chem. Ecol.* **12**, 1285–1293 (1986).
- Heinze, J., Oberstadt, B., Tentschert, J., Hölldobler, B. & Bestmann, H. J. Colony specificity of Dufour gland secretions in a functionally monogynous ant. *Chemoecology* **8**, 169–174 (1998).
- Gobin, B., Billen, J. & Peeters, C. Policing behaviour towards virgin egg layers in a polygynous ponerine ant. *Anim. Behav.* **58**, 1117–1122 (1999).
- Liebig, J., Peeters, C. & Hölldobler, B. Worker policing limits the number of reproductives in a ponerine ant. *Proc. R. Soc. Lond. B* **266**, 1865–1870 (1999).
- Monnin, T. & Ratnieks, F. L. W. Policing in queenless ponerine ants. *Behav. Ecol. Sociobiol.* **50**, 97–108 (2001).
- Monnin, T., Malosse, C. & Peeters, C. Solid phase microextraction and cuticular hydrocarbon differences related to reproductive activity in the queenless ant *Dinoponera quadricaps*. *J. Chem. Ecol.* **24**, 473–490 (1998).
- Liebig, J., Peeters, C., Oldham, N. J., Markstädter, C. & Hölldobler, B. Are variations in cuticular hydrocarbons of queens and workers a reliable signal of fertility in the ant *Harpegnathos saltator*? *Proc. Natl Acad. Sci. USA* **97**, 4124–4131 (2000).
- Cuvillier-Hot, V., Cobb, M., Malosse, C. & Peeters, C. Sex, age and ovarian activity affect cuticular hydrocarbons in *Diacamma ceylonense*, a queenless ant. *J. Insect Physiol.* **47**, 485–493 (2001).
- Stacey, P. B. & Koenig, W. D. *Cooperative Breeding in Birds. Long Term Studies of Ecology and Behaviour* (Cambridge Univ. Press, Cambridge, 1990).
- Solomon, N. G. & French, J. A. *Cooperative Breeding in Mammals* (Cambridge Univ. Press, Cambridge, 1997).
- Ratnieks, F. L. W. Reproductive harmony via mutual policing by workers in eusocial hymenoptera. *Am. Nat.* **132**, 217–236 (1988).

26. Ratnieks, F. L. W. & Visscher, P. K. Worker policing in the honeybee. *Nature* **342**, 796–797 (1989).
27. Barron, A., Oldroyd, B. P. & Ratnieks, F. L. W. Worker reproduction in honey-bees (*Apis*) and the anarchic syndrome: a review. *Behav. Ecol. Sociobiol.* **50**, 199–208 (2001).
28. Foster, K. R. & Ratnieks, F. L. W. Social insects: Facultative worker policing in a wasp. *Nature* **407**, 692–693 (2000).
29. Foster, K. R. & Ratnieks, F. L. W. Convergent evolution of worker policing by egg eating in the honeybee and common wasp. *Proc. R. Soc. Lond. B* **268**, 169–174 (2001).
30. Peeters, C., Monnin, T. & Malosse, C. Cuticular hydrocarbons correlated with reproductive status in a queenless ant. *Proc. R. Soc. Lond. B* **266**, 1323–1327 (1999).

Acknowledgements

We thank IBAMA for permission to collect *Dinoponera* in Brazil, and T. Clutton-Brock, M. Cobb, K. Foster, A. Hefetz, L. Keller, J. Liebig and C. Peeters for comments. T.M. was funded by a Marie Curie individual fellowship from the European Union programme Training and Mobility of Researchers.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.M.

(e-mail: Thibaud.Monnin@snv.jussieu.fr).

An ultra-sparse code underlies the generation of neural sequences in a songbird

Richard H. R. Hahnloser*†‡, Alexay A. Kozhevnikov*‡ & Michale S. Fee*

* Biological Computation Research Department, Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

† Howard Hughes Medical Institute, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

‡ These authors contributed equally to this work

Sequences of motor activity are encoded in many vertebrate brains by complex spatio-temporal patterns of neural activity; however, the neural circuit mechanisms underlying the generation of these pre-motor patterns are poorly understood. In songbirds, one prominent site of pre-motor activity is the fore-brain robust nucleus of the archistriatum (RA), which generates stereotyped sequences of spike bursts during song¹ and recapitulates these sequences during sleep². We show that the stereotyped sequences in RA are driven from nucleus HVC (high vocal centre), the principal pre-motor input to RA^{3,4}. Recordings of identified HVC neurons in sleeping and singing birds show that individual HVC neurons projecting onto RA neurons produce bursts sparsely, at a single, precise time during the RA sequence. These HVC neurons burst sequentially with respect to one another. We suggest that at each time in the RA sequence, the ensemble of active RA neurons is driven by a subpopulation of RA-projecting HVC neurons that is active only at that time. As a population, these HVC neurons may form an explicit representation of time in the sequence. Such a sparse representation, a temporal analogue of the ‘grandmother cell’⁵ concept for object recognition, eliminates the problem of temporal interference during sequence generation and learning attributed to more distributed representations^{6,7}.

Songbirds produce highly stereotyped, learned vocalizations^{8,9}. Zebra finch (*Taeniopygia guttata*) song consists of a complex pattern of sounds with spectral and temporal modulation over a wide range of timescales¹⁰. A basic acoustic element is the song syllable, which may itself be composed of a complex sequence of sounds varying on a 10-ms timescale, or even less¹¹. Several distinct song syllables are organized into a single, repeated pattern of about 1 s in duration,

called a song motif. Two pre-motor nuclei have been identified for their importance in song generation: nucleus RA and nucleus HVC¹². Premotor HVC neurons project onto RA neurons, which in turn project with a myotopic mapping onto motor neurons of the vocal organ¹³, and to respiratory brain areas¹⁴. During singing, RA neurons generate a highly stereotyped, complex sequence of action potential bursts, each precisely correlated to the song vocalization on a submillisecond timescale^{1,15}. The average burst duration is roughly 10 ms, and each RA neuron generates a unique pattern of roughly ten bursts per song motif, such that on average 12% of RA neurons are active at any time (A. Leonardo, and M.S.F., unpublished data) (Fig. 1a).

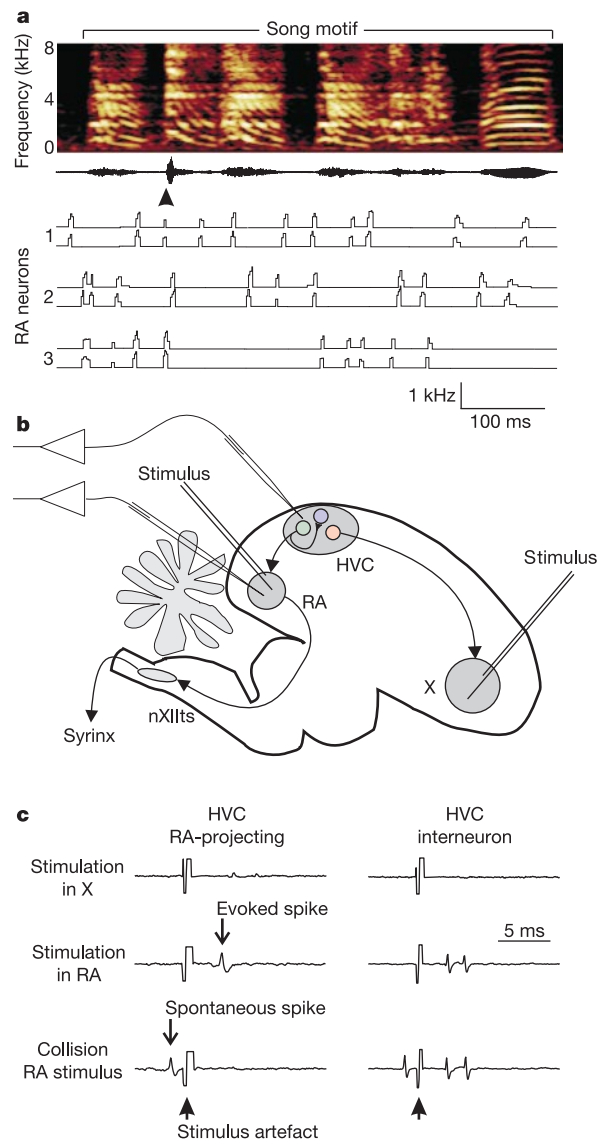


Figure 1 RA sequences and identification of HVC neurons. **a**, Neurons in nucleus RA generate complex sequences of brief action potential bursts during song vocalizations. Spectrogram (top) and acoustic signal of the song motif, and plots of instantaneous firing rate (bottom) of song-related spike activity in three different RA neurons recorded in one zebra finch. Neural activity is aligned using the onset of the second syllable of each motif (arrowhead). Two renditions are displayed for each neuron. **b**, Single-unit recordings were made in pre-motor nuclei HVC and RA. HVC neurons were antidromically identified by electrical stimulation in RA and area X. RA projects to vocal motor neurons in the nucleus of the twelfth nerve (nXlts). **c**, RA-projecting neurons and putative interneurons could be activated from RA but not from area X. Stimulation in RA, triggered by spontaneous spikes, resulted in spike collision for RA-projecting neurons but not for interneurons.

Here, we avoid the question of how RA activity is translated into sound, and simply ask how pre-motor burst patterns in RA are generated. Previous studies have suggested that the syllable order and tempo of the motif are generated by a network that resides above RA, and includes HVC^{12,16}, and that an HVC neural code for syllables is transformed into a code for shorter acoustic elements through the projection of HVC onto RA^{1,17}. To re-examine these issues, we have characterized the role of inputs to RA from pre-motor nucleus HVC.

HVC contains at least three classes of neurons: neurons that project to the RA, neurons that project to area X, and interneurons^{18,19}. We have identified HVC neuron classes by antidromic activation²⁰ from RA and from area X (Fig. 1b, c). Chronic single-neuron recordings were made from identified neurons of all three classes. Antidromically identified RA-projecting HVC neurons ($HVC_{(RA)}$) ($n = 16$, three birds) were completely inactive in awake, non-singing birds (<0.001 spikes s^{-1}), and burst extremely sparsely during vocalizations, generating at most a single burst per song motif (Fig. 2a). $HVC_{(RA)}$ bursts had a duration of 6.1 ± 2 ms, and comprised 4.5 ± 2 spikes at a firing rate of $613 \pm 210 s^{-1}$ (ranges are ± 1 s.d. unless specified otherwise). $HVC_{(RA)}$ bursts were highly stereotyped, tightly time-locked to the song motif (0.66 ± 0.14 ms r.m.s. jitter), and occurred reliably on every rendition of the motif (Fig. 2b). Thus, on a millisecond timescale, $HVC_{(RA)}$ bursts were maximally correlated to the vocalization. Different $HVC_{(RA)}$ neurons tended to burst at different times in the song, with no obvious timing relation to the onset or offset of song syllables. Three identified $HVC_{(RA)}$ neurons generated no bursts during the song, but produced a single burst during call vocalizations. HVC neurons projecting to area X also burst sparsely during singing (0–5 bursts per motif, $n = 30$; data not shown). In contrast to projection neurons, putative HVC interneurons ($n = 31$), most of which were spontaneously active in the non-singing bird (11 ± 7 spikes s^{-1}), produced high rates of spiking and

bursting activity throughout song and call vocalizations (Fig. 2b). The firing patterns of putative HVC interneurons were similar to those of unidentified neurons found in previous studies of HVC in the singing bird¹.

Previous observations have shown that sleep-related spike and burst patterns in nucleus RA can closely recapitulate those generated during singing², suggesting that a common neural mechanism may underlie the generation of song- and sleep-related RA burst patterns. A more detailed understanding of the role of HVC in generating sleep-related activity in RA may provide a hint as to the interaction of these two nuclei during singing. We next examined the firing patterns of RA neurons and identified HVC neurons using a new, sleeping-bird preparation where the head of the bird is fixed, permitting simultaneous single-unit recordings in multiple brain areas and pharmacological manipulation, which are not currently possible in the singing bird.

Similar to the situation in the singing bird, $HVC_{(RA)}$ neurons burst sparsely during sleep (0.06 ± 0.05 bursts s^{-1} , $n = 116$, 27 birds). Paired recordings in RA and HVC (Fig. 3a) neurons showed that $HVC_{(RA)}$ neurons fired 13 ± 3 times fewer bursts in the sleeping bird than did RA neurons ($n = 53$ pairs). The bursts had properties similar to those observed during singing: duration of bursts during sleep in RA and $HVC_{(RA)}$ neurons were 11.5 ± 3.5 ms and 6.5 ± 1.8 ms, respectively. Bursts of $HVC_{(RA)}$ neurons during sleep comprised 3.2 ± 0.8 spikes per burst, and had an average firing rate of $347 \pm 81 s^{-1}$. The relationship between $HVC_{(RA)}$ bursts and RA bursts is readily seen in raster plots of RA spike trains aligned in time to the onset of bursts in $HVC_{(RA)}$ neurons (Fig. 3b, c). RA neurons reliably showed a pattern of bursts locked to the $HVC_{(RA)}$ bursts ($n = 45$ of 53 pairs). Furthermore, multiple RA neurons recorded sequentially with a single $HVC_{(RA)}$ neuron ($n = 3$) show that different RA neurons generate different patterns of bursts, as is the case during singing. The relation between $HVC_{(RA)}$ and RA spike trains was quantified using a correlation

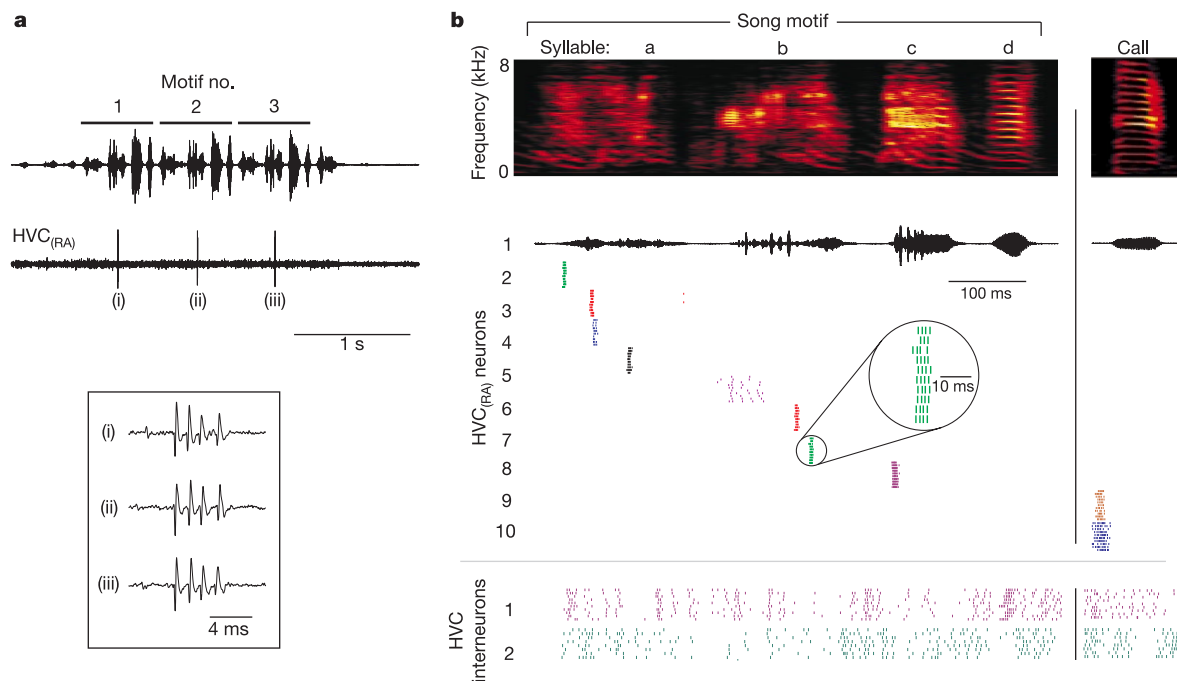


Figure 2 Spiking activity of identified HVC neurons during singing. **a**, Extracellular record of an RA-projecting HVC ($HVC_{(RA)}$) neuron (bottom), with the simultaneously recorded vocalization (top). The $HVC_{(RA)}$ neuron generates a single burst during each of three motif renditions. **b**, Spike raster plot of ten $HVC_{(RA)}$ neurons and two HVC interneurons recorded in one bird during singing (left) and call vocalizations (right). Each row of tick marks shows

spikes generated during one rendition of the song or call; roughly ten renditions are shown for each neuron. Neural activity is aligned by the acoustic onset of the nearest syllable. $HVC_{(RA)}$ neurons burst reliably at a single precise time in the song or call; however, HVC interneurons spike or burst densely throughout the vocalizations.

function, $K(\tau)$, that is conditional on spikes in the $HVC_{(RA)}$ neuron ($-1 < K(\tau) < 1$) (ref. 21) (see Methods). Unlike the standard cross-correlation function, $K(\tau)$ is not sensitive to large differences in firing rate of the two spike trains. The peaks in the conditional correlation function in RA– $HVC_{(RA)}$ pairs were large; the largest peak correlation for each pair was on average $\bar{K} = 0.62$ ($n = 45$ pairs). Multiple significant peaks in $K(\tau)$, corresponding to reliable bursts in the RA sequence, were found over a wide range of time lags (± 400 ms full range, ± 130 ms s.d.), with weaker correlations at larger time lags (Fig. 3d). Significant peaks were approximately symmetrically distributed around zero time lag; that is, bursts in the RA sequence were observed both before and after the $HVC_{(RA)}$ bursts.

Paired recordings of $HVC_{(RA)}$ neurons reveal that these neurons generate a sparse sequence during sleep (Fig. 3e), consistent with our observations in the singing bird. Nine $HVC_{(RA)}$ pairs (out of 26 pairs, 7 birds) showed a statistically significant conditional correlation; in only one of these pairs did the neurons burst simultaneously. The time lags of the peaks were distributed over several hundred milliseconds (Fig. 3f), showing that the population of $HVC_{(RA)}$ neurons forms a sequence over a timescale similar to that of the sleep sequences in RA neurons. $HVC_{(RA)}$ pairs recorded on a single electrode revealed that neighbouring $HVC_{(RA)}$ neurons did not burst at the same time ($n = 3$ in the sleeping bird, $n = 1$ in the singing bird), suggesting that there may be no spatial organization to the burst times in HVC.

In the singing bird, $HVC_{(RA)}$ neurons burst at most once per motif, thus spike train autocorrelations of these neurons do not exhibit any peaks at time lags greater than 10 ms and less than the motif duration. Similarly, the autocorrelations of $HVC_{(RA)}$ neurons recorded during sleep did not show significant peaks at time lags greater than 10 ms (134 of 135 neurons). Furthermore, during singing, pairs of $HVC_{(RA)}$ neurons are highly correlated at a single time lag corresponding to the interval between their burst times in the motif (as inferred from their correlation to the vocalization, see Fig. 2). Similarly, during sleep, correlated $HVC_{(RA)}$ pairs show only a single significant peak in the conditional correlation function (for 8 of 9 correlated pairs, Fig. 3f).

Do action potential bursts in $HVC_{(RA)}$ neurons drive bursts in RA neurons? Reversible lesions in HVC demonstrate that spiking activity in HVC is necessary for the generation of sleep sequences in RA. In three birds, microinjection into HVC of lidocaine, a sodium channel blocker that suppresses spiking, completely and reversibly abolished sleep bursts in RA. Total recovery was observed 2–3 min after injection. Injections in HVC had no effect on the spontaneous 10–20-Hz regular spiking in RA neurons, confirming that lidocaine did not diffuse into RA. Control injections of saline had no effect on sleep bursts ($n = 3$ birds). Injections of lidocaine into forebrain nucleus LMAN²², a non-pre-motor input to RA, did not affect the generation of bursts during sleep ($n = 2$ birds).

Evidence that bursts in RA are driven in a feed-forward manner by preceding activity in the HVC is also provided by observations of

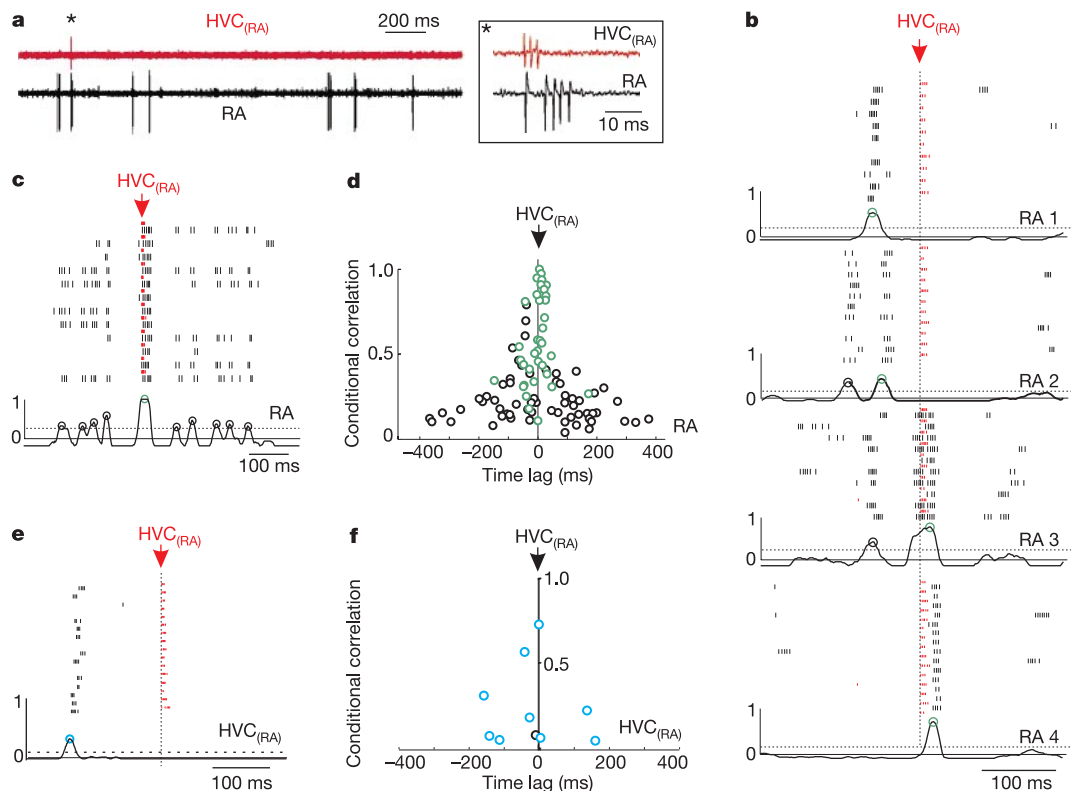


Figure 3 Relationship between burst patterns in a sleeping bird of RA neurons and $HVC_{(RA)}$ neurons. **a**, Extracellular signals from an $HVC_{(RA)}$ neuron recorded simultaneously with an RA neuron. $HVC_{(RA)}$ neurons burst sparsely, approximately 13 times less often than RA neurons. **b**, Spike raster plot of simultaneously recorded pairs of $HVC_{(RA)}$ and RA neurons. Shown is one $HVC_{(RA)}$ neuron recorded sequentially with four different RA neurons. All $HVC_{(RA)}$ bursts (short red rasters) are aligned at the centre of the plot. Corresponding RA spikes (long black rasters) are shown below each $HVC_{(RA)}$ burst. Below each raster is shown the conditional correlation function. **c**, Another example of an RA burst sequence in a sleeping bird (without melatonin). Green circles show the largest statistically significant

correlation peak; black circles show statistically significant secondary peaks. The horizontal dashed line is the 95% confidence level, as assessed by Monte-Carlo estimation. **d**, Distribution of peak correlation and time lag of statistically significant main peaks (green) and secondary peaks (black). Positive time indicates that the RA neuron follows the $HVC_{(RA)}$ neuron. **e**, Raster plot of simultaneously recorded $HVC_{(RA)}$ neurons, with conditional correlation function below. **f**, Peak values and associated time lags of the conditional correlation functions for $HVC_{(RA)}$ pairs (blue circles show significant main peaks ($P < 0.05$), the black circle shows the only secondary peak).

putative HVC interneurons ($HVC_{(i)}$) in the sleeping bird. Paired recordings showed that the burst density of $HVC_{(i)}$ neurons was roughly three times the burst density of simultaneously recorded RA neurons ($n = 35$ pairs) and 45 times that of $HVC_{(RA)}$ neurons ($n = 22$ pairs). Bursts of $HVC_{(RA)}$ neurons were strongly synchronized with bursts of $HVC_{(i)}$ neurons (Fig. 4a; average correlation $\bar{K} = 0.73$, average latency $\tau = 2.5 \pm 3.0$ ms ($\pm$ s.e.m.)). Similarly, most bursts of RA neurons strongly correlated with preceding bursts of $HVC_{(i)}$ neurons (Fig. 4a, c; average correlation $\bar{K} = 0.51$, average latency $\tau = 4.5 \pm 0.4$ ms ($\pm$ s.e.m.)). The average spike latencies of RA and $HVC_{(RA)}$ neurons relative to $HVC_{(i)}$ neurons are consistent with the measured distribution of antidromic $HVC_{(RA)}$ spike latencies (3.5 ± 2.5 ms ($\pm$ s.d.)). Because there is no known direct anatomical projection from RA to HVC, these results, together with the HVC lesion experiments, imply a short timescale (~ 4.5 ms) and causal relationship between $HVC_{(RA)}$ bursts and RA bursts.

The strong correlation between $HVC_{(RA)}$ and $HVC_{(i)}$ neurons suggests that activity in single $HVC_{(i)}$ neurons during sleep is representative of the population activity of $HVC_{(RA)}$ neurons. Consistent with this idea, all interneurons we observed were strongly activated by near-threshold antidromic stimulation of $HVC_{(RA)}$ neurons; additionally HVC interneurons were highly synchronized as a population during sleep (Fig. 4b). In paired recordings, nearly every sleep burst in one $HVC_{(i)}$ neuron occurred

simultaneously with a burst in the other $HVC_{(i)}$ neuron ($\bar{K} = 0.78$, $n = 19$ pairs; Fig. 4c). Thus, the fact that bursts in $HVC_{(i)}$ neurons precede most sleep bursts in RA suggests that feed-forward activation from HVC dominates the generation of sleep sequences in RA. The feed-forward contribution, that is, the fraction of RA bursts that are driven from HVC, is given to a first approximation by the ratio of the RA– $HVC_{(i)}$ conditional correlation to the $HVC_{(RA)}$ – $HVC_{(i)}$ conditional correlation. (The $HVC_{(RA)}$ – $HVC_{(i)}$ conditional correlation is a measure of how well the $HVC_{(i)}$ neurons represent activity in the $HVC_{(RA)}$ population, and is therefore used as a normalization.) Thus, our estimate of the HVC feed-forward contribution to RA sleep sequences is 70%.

To what extent does this and other observations in the sleeping bird inform our understanding of the generation of RA song sequences? Are sleep sequences a transient activation of the same neural mechanisms that underlie the generation of sequences during singing? Previous studies indicate that during sleep, individual RA neurons can generate the same unique sequence of spikes and bursts as generated by that neuron during singing². Our data do not allow a comparison of individual RA– $HVC_{(RA)}$ pair correlations in the sleeping and singing bird; however, we can compare RA and HVC correlations on a population basis in these two states. Although average correlations are weaker for sleep sequences, they share the following features with song sequences: (1) $HVC_{(RA)}$ neurons, as a population, generate sparse sequences; (2) most RA neurons generate a unique pattern of bursts in relation to a given $HVC_{(RA)}$ neuron burst; (3) the burst widths, and the ratio of burst densities of RA and $HVC_{(RA)}$ neurons are similar in both behavioural states; and (4) HVC interneurons fire or burst densely during both song and sleep sequences. In conclusion, the relationship between song and sleep sequences of individual neurons is unresolved, but our results suggest that the generation of RA song sequences may also be dominated by a feed-forward drive from HVC; that is, most of the RA bursts during singing are driven directly by bursts in $HVC_{(RA)}$ neurons.

Our finding that $HVC_{(RA)}$ neurons produce a brief burst at a single time in only one of the repeating vocal elements (that is, motif or call), and that, as a population, $HVC_{(RA)}$ neurons are probably active throughout the associated RA sequences, suggests that these neurons code for some transient event or 'state' in the sequence, rather than longer timescale structures such as song syllables, as was previously believed. We propose that $HVC_{(RA)}$ neurons code for a unique time within the RA sequence; that is, that they are active at one and only one time point in the RA sequence, regardless of motor or vocal content. This does not imply that $HVC_{(RA)}$ neurons can never burst more than once in a song motif: an $HVC_{(RA)}$ neuron active within a syllable identically repeated in a motif will probably burst during each repeat of the syllable, as suggested by the precisely repeated RA sequence associated with identical syllables within a motif¹ (A. Leonardo and M.S.F., unpublished observations). Given the precise relationship between $HVC_{(RA)}$ bursts and RA burst patterns however, it seems unlikely that $HVC_{(RA)}$ neurons will burst several times within a motif that does not have repeated RA, and therefore acoustic, sequences.

In most neural models of sequence generation, a sequence of activity is stored by specific excitatory and inhibitory connections between neurons within the network, such that the network advances autonomously from one state to the next^{23,24}. Such single-layer recurrent networks suffer from the problem of temporal interference due to spatial overlap of different states in the sequence, and consequently must operate with very low average activity^{6,7}, cannot produce the same state at multiple times during the sequence, or are difficult to train²⁵. The previous theoretical work showed that a sparse representation, such as we observe in HVC, resolves these interference problems. Our data indicate that the dynamics that underlie RA sequence generation occur not in RA but in another network (for example, the HVC and higher) that

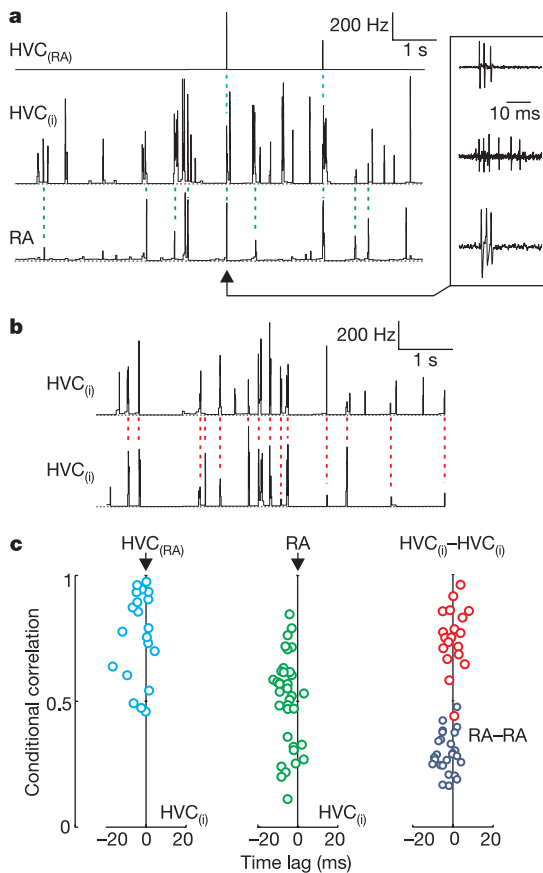


Figure 4 Global synchrony of HVC interneurons. **a**, Simultaneous single-unit triplet recording, during sleep, of an $HVC_{(RA)}$ neuron, a putative HVC interneuron ($HVC_{(i)}$) and an RA neuron. Plotted are instantaneous firing rate functions. RA bursts and $HVC_{(RA)}$ bursts are strongly synchronized with interneuron bursts. **b**, Instantaneous firing rate function for a simultaneously recorded pair of $HVC_{(i)}$ neurons, which show highly synchronized bursts. **c**, Scatter plot of conditional correlation versus peak times of 20 $HVC_{(RA)}$ – $HVC_{(i)}$ pairs, 35 RA– $HVC_{(i)}$ pairs, 19 $HVC_{(i)}$ – $HVC_{(i)}$ pairs, and 27 RA–RA pairs. For the left two plots, the time lag is plotted relative to the bursts of the $HVC_{(RA)}$ neuron and the RA neuron, respectively.

incorporates a canonical sparse representation. The canonical sequence in HVC is translated into a complex sequence in RA by a set of synaptic weights from $HVC_{(RA)}$ to RA neurons; thus RA can, in principle, produce arbitrary sequences with high average levels of activity and large state overlaps.

Put in terms of songbird behaviour, the motor actions underlying song unfold when the 'time code' in HVC acts on the 'muscle map' in RA. The pattern of activation in RA depends on the synaptic connections from $HVC_{(RA)}$ neurons onto RA neurons, suggesting that it is the pattern of $HVC_{(RA)}$ –RA synapses that contains the 'score' of the bird's song. Our results are thus consistent with the idea that plasticity in RA mediates song learning^{26–28}, and suggest that the site of this plasticity is the $HVC_{(RA)}$ –RA synapse. Of course we cannot exclude the possibility that song learning may also involve plasticity in the order in which $HVC_{(RA)}$ neurons burst.

The ultimately sparse, or unary, representation in HVC may simplify sequence learning as well: errors in the song at a particular time in the sequence can be corrected, through auditory feedback, by altering only those $HVC_{(RA)}$ –RA synapses that were active at the time of, or immediately preceding, the error. In a dense representation, in which $HVC_{(RA)}$ neurons would burst many times, changes in synaptic connectivity intended to reduce song errors at one time would affect song output at many times, resulting in a difficult learning process. However, with a sparse representation in HVC, changes in the $HVC_{(RA)}$ –RA synapses during learning will modify the RA pattern, and thus the song, only at the time of the error, without introducing unwanted changes in the song at other times. □

Methods

Electrophysiology

Subjects were adult male zebra finches. The care and experimental manipulation of the animals was carried out in accordance with guidelines of the National Institute of Health and have been reviewed and approved by the local Institutional Animal Care and Use Committee. Before surgery, anaesthesia was produced with 1–3% isoflurane in oxygen. For experiments on sleeping birds, a thin stainless-steel plate was mounted to the skull with dental acrylic, and a window was made in the inner and outer bone leaflet over nuclei RA, HVC and area X of the right hemisphere using stereotaxic coordinates. A small hole (~200 µm) was made in the dura over each area and the exposed brain was protected with 2% low melting agarose in saline. Wound margins were treated with lidocaine gel. The animal was placed in a small foam restraint and placed in the recording apparatus without further anaesthesia. Birds that did not sleep spontaneously were administered subcutaneously 10 µg melatonin (Sigma) in phosphate buffered saline²⁹. Additional details on assessing sleep state are described in the Supplementary Information. For experiments in the singing bird, single-unit recordings were made with a three-channel motorized microdrive, using previously described techniques³⁰. For chronic and head-fixed HVC recordings, stimulating electrodes were implanted in RA and area X for antidromic activation of HVC. Data were obtained from 41 zebra finches.

In the sleeping bird, simultaneous paired and triplet single-neuron recordings were made in nucleus RA and HVC. In the singing bird, single neurons were recorded in HVC or in RA. High signal-to-noise (greater than 5:1) recordings were made with sharp glass and platinum/tungsten microelectrodes. All spike trains were confirmed to be single-unit by complete suppression of the spike autocorrelation functions at times less than 1 ms. Antidromic identification of HVC neuron type was based on spike collision and spike latency variability²⁰, as described in Supplementary Information.

After each experiment, the animal was killed by intraperitoneal injection of 20% urethane (300 µl). The brain was removed for histological examination of unstained slices to verify the location of stimulating and recording electrodes, and drug injection sites.

Data analysis

In Figs 1 and 4, neural activities are represented as instantaneous firing rates, $R(t)$, defined at each time as the inverse of the enclosed interspike interval, $R(t) = 1/(t_{i+1} - t_i)$, for $t_i < t \leq t_{i+1}$, where t_i is the time of the i th spike. A burst is defined as the interval over which the instantaneous firing rate exceeds 100 Hz. The burst rate is defined as the number of events per unit time in which the instantaneous firing rate crosses the 100-Hz threshold from below.

We use a measure of correlation between two simultaneously recorded spike trains, A and B, which is conditional on the spike train with fewer spikes. This measure is insensitive to differences in average firing rate. The conditional correlation function, $K(\tau)$, represents the excess fraction of spikes in train A that fall within a window $(\tau - \Delta t, \tau + \Delta t)$ of a spike in train B, where Δt is the half-width of the window. The spike times of A and B are denoted by t_i^A ($1 \leq i \leq N^A$) and t_j^B ($1 \leq j \leq N^B$), respectively, where A has a smaller number of spikes than B ($N^A < N^B$). In calculating the conditional correlation, we define the conditional spike probability, $c(\tau)$ (ref. 21), as the fraction of spikes in A that are within Δt of at least one spike in B. $c(\tau)$ is calculated as a function of an arbitrary time shift, τ ,

between the two spike trains. Thus,

$$c(\tau) = \frac{1}{N^A} \sum_i \Theta[\Delta t - \min_j (|t_i^A - t_j^B - \tau|)]$$

where $\Theta[x]$ is the Heavyside function: $\Theta[x] = 0$ if $x < 0$ and $\Theta[x] = 1$ otherwise. Note that $0 \leq c(\tau) \leq 1$.

The conditional correlation is defined as: $K(\tau) = (c(\tau) - \bar{c})/(\bar{c}(1 - \bar{c}))$, for $\bar{c} \leq 0.5$ where $\bar{c} = \langle c(\tau) \rangle_\tau$ denotes the average conditional spike probability calculated on randomly shifted spike trains, as described below. The conditional correlation characterizes the two spike trains as being somewhere between anti-correlated ($K = -1$) and strongly correlated ($K = 1$), and obeys the inequalities $-1 \leq -\bar{c}/(1 - \bar{c}) \leq K \leq 1$. Apart from normalization, the shape of the conditional correlation function is qualitatively similar to the standard cross-correlation; in particular, the time lags at the peaks are nearly identical. We use a coincidence window of $\Delta t = 5$ ms, half of the typical burst width and half of the defined maximum interspike interval within bursts. The correlation is calculated for time lags over a ± 1 -s window.

From $K(\tau)$, we extract all peak values and their time lags τ . The significance of these peaks is assessed in relation to randomly shifted spike trains: both spike trains were divided into 500-ms windows, each of which was given a cyclic time shift randomly chosen in the range of 0 to 500 ms. (Sleep bursts tended to occur in epochs of 1 to 2 s; thus, shifting spike trains over intervals longer than 500 ms produces a weaker test of significance.) Peak values greater than zero were determined from 300 such randomized spike trains, typically yielding several thousand peaks. The distribution of peak values was used to find the 95% confidence level. Additional details on the computation of conditional correlations is described in the Supplementary Information.

Received 26 February; accepted 21 June 2002; doi:10.1038/nature00974.

- Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273**, 1871–1875 (1996).
- Dave, A. S. & Margoliash, D. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* **290**, 812–816 (2000).
- Nottebohm, F., Kelley, D. B. & Paton, J. A. Connections of vocal control nuclei in the canary telencephalon. *J. Comp. Neurol.* **207**, 344–357 (1982).
- Scharff, C. & Nottebohm, F. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J. Neurosci.* **11**, 2896–2913 (1991).
- Barlow, H. *The Cognitive Neurosciences* 415–435 (MIT Press, Cambridge, Massachusetts, 1995).
- Hermann, M., Hertz, J. & Prugel-Bennet, A. Analysis of synfire chains. *Network: Comput. Neural Syst.* **6**(3), 403–414 (1995).
- Willshaw, D. J., Buneman, O. P. & Longuet-Higgins, H. C. Non-holographic associative memory. *Nature* **222**, 960–962 (1969).
- Konishi, M. The role of auditory feedback in the control of vocalizations in the white-crowned sparrow. *Z. Tierpsychol.* **22**, 770–783 (1965).
- Konishi, M. Birdsong: from behavior to neuron. *Ann. Rev. Neurosci.* **8**, 125–170 (1985).
- Immelmann, K. In *Bird Vocalizations* (ed. Hinde, R. A.) 61–74 (Cambridge Univ. Press, New York, 1969).
- Fee, M. S., Shraiman, B., Pesaran, B. & Mitra, P. P. The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature* **395**, 67–71 (1998).
- Vu, E. T., Mazurek, M. E. & Kuo, Y. Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci.* **14**, 6924–6934 (1994).
- Vicario, D. S. Organization of the zebra finch song control system: II. Functional organization of outputs from nucleus robustus archistriatalis. *J. Comp. Neurol.* **309**, 486–494 (1991).
- Wild, J. M. Descending projections of the songbird nucleus robustus archistriatalis. *J. Comp. Neurol.* **338**, 225–241 (1993).
- Chi, Z. & Margoliash, D. Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron* **32**, 899–910 (2001).
- Doya, K. & Sejnowski, T. J. In *Central Auditory Processing and Neural Modeling* (eds Poon, P. W. F. & Brugge, J. F.) 77–88 (Plenum, New York, 1998).
- Troyer, T. W. & Doupe, A. J. An associational model of birdsong sensorimotor learning II. Temporal hierarchies and the learning of song sequence. *J. Neurophysiol.* **84**, 1224–1239 (2000).
- Mooney, R. Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the zebra finch. *J. Neurosci.* **20**, 5420–5436 (2000).
- Dutar, P., Vu, H. M. & Perkel, D. J. Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in nucleus HVC of the adult zebra finch. *J. Neurophysiol.* **80**, 1828–1838 (1998).
- Swadlow, H. Neocortical efferent neurons with very slowly conducting axons: strategies for reliable antidromic identification. *J. Neurosci. Meth.* **79**, 131–141 (1998).
- Brillinger, D. R., Bryant, H. L. & Segundo, J. P. Identification of synaptic interactions. *Biol. Cybernetics* **22**, 213–228 (1976).
- Doupe, A. J. A neural circuit specialized for vocal learning. *Curr. Opin. Neurobiol.* **11**, 104–111 (1993).
- Amari, S.-I. Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Trans. Comput.* **11**, 1197–1206 (1972).
- Kleinfeld, D. & Sompolinsky, H. In *Methods in Neuronal Modeling* (eds Koch, C. & Segev, I.) 195–246 (MIT Press, Cambridge, Massachusetts, 1989).
- Pearlmutter, B. A. Learning state space trajectories in recurrent neural networks. *Neural Computat.* **1**, 263–269 (1989).
- Bottjer, S. W., Halsema, K. A., Brown, S. A. & Miesner, E. A. Axonal connections of a forebrain nucleus involved with vocal learning in zebra finches. *J. Comp. Neurol.* **279**, 312–326 (1989).
- Mooney, R. Synaptic basis for developmental plasticity in a birdsong nucleus. *J. Neurosci.* **12**, 2464–2477 (1992).
- Brainard, M. S. & Doupe, A. J. Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* **404**, 762–766 (2000).
- Phillips, N. H. & Berger, R. J. Melatonin infusions restore sleep suppressed by continuous bright light in pigeons. *Neurosci. Lett.* **145**, 217–220 (1992).
- Fee, M. S. & Leonardo, A. Miniature motorized microdrive and commutator system for chronic neural recordings in small animals. *J. Neurosci. Meth.* **112**, 83–94 (2001).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We acknowledge discussions with W. Denk, D. Lee, I. Nebel and S. Seung. We also thank F. Nottelbohm for comments on the manuscript. Recordings of RA neurons in the singing bird were carried out in collaboration with A. Leonardo. This work was supported in part by the National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S.F. (e-mail: fee@lucent.com).

Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes

Karina Del Punta*, Trese Leinders-Zufall†, Ivan Rodriguez*‡, David Jukam*, Charles J. Wysocki§, Sonoko Ogawa*, Frank Zufall† & Peter Mombaerts*

* The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

† Department of Anatomy and Neurobiology and Program in Neuroscience, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA

§ Monell Chemical Senses Center, 3500 Market Street, Philadelphia, Pennsylvania 19104, USA

The mammalian vomeronasal organ (VNO), a part of the olfactory system, detects pheromones—chemical signals that modulate social and reproductive behaviours^{1,2}. But the molecular receptors in the VNO that detect these chemosensory stimuli remain undefined. Candidate pheromone receptors are encoded by two distinct and complex superfamilies of genes, *V1r* and *V2r* (refs 3 and 4), which code for receptors with seven transmembrane domains. These genes are selectively expressed in sensory neurons of the VNO. However, there is at present no functional evidence for a role of these genes in pheromone responses. Here, using chromosome engineering technology⁵, we delete in the germ line of mice a ~600-kilobase genomic region that contains a cluster of 16 intact *V1r* genes⁶. These genes comprise two of the 12 described *V1r* gene families⁷, and represent ~12% of the *V1r* repertoire. The mutant mice display deficits in a subset of VNO-dependent behaviours: the expression of male sexual behaviour and maternal aggression is substantially altered. Electrophysiologically, the epithelium of the VNO of such mice does not respond detectably to specific pheromonal ligands. The behavioural impairment and chemosensory deficit support a role of *V1r* receptors as pheromone receptors.

Vomeronasal sensory neurons (VSNs) in the mammalian VNO are thought to be specialized in the detection of pheromones^{8,9}, although their chemoreceptive abilities also extend to other types of ligands¹⁰. Much of our knowledge about mammalian VNO function has been obtained by describing the behavioural consequences of surgical lesions. Animals with complete anatomical removal of the VNO exhibit a range of deficits in social and reproductive behaviours, including the emission of male ultrasound vocalizations to females, intermale aggression, male sexual behaviour, and

maternal aggression¹¹. Pheromones, present in bodily secretions such as urine^{1,2}, elicit these VNO-dependent behaviours. *V1r* (ref. 3) and *V2r* (ref. 4) genes are proposed to encode candidate chemosensory receptors, specifically pheromone receptors, on the basis of predicted protein structure, patterns of expression, and genetic complexity. However, no ligands, natural or synthetic, are known for specific *V1r* or *V2r* receptors and there is no functional evidence implicating these receptors in pheromonal behaviours.

To determine if *V1r* receptors are involved in behavioural and electrophysiological responses to pheromones, we deleted a cluster of *V1r* genes in the germ line of mice. We⁶ and others¹² have characterized the genomic organization of a cluster of *V1r* genes on chromosome 6. A ~600-kilobase (kb) region comprises 23 *V1r* genes of which 16 have an intact open reading frame (ORF); the remaining 7 *V1r* genes are pseudogenes, as their ORFs are disrupted. The mouse *V1r* superfamily⁷ consists of ≥137 genes with an intact ORF that can be grouped into 12 phylogenetically highly isolated families (*V1ra-l*) (Fig. 1a). Our cluster represents ~12% of the potentially functional *V1r* repertoire, and contains most members of two *V1r* families, *V1ra* and *V1rb* (Fig. 1a, b), but no other genes^{6,12}. To delete the gene cluster we used the technique of chromosome engineering⁵, designed to excise large regions of genomic DNA via gene targeting and Cre-*loxP* mediated site-specific recombination in embryonic stem cells (Fig. 1b). The resulting homozygous mice, termed $\Delta V1rab\Delta$ mice, harbour a ~600-kb genomic deletion and lack most *V1ra* and *V1rb* genes, as evidenced by Southern blot hybridization with a *V1ra/b* probe (Fig. 1c). One intact family member, *V1rb10*, located on the X chromosome⁶, is not included in the deletion. *In situ* hybridization of the vomeronasal epithelium confirmed the expected absence of messenger RNA for the deleted *V1r* genes in the mutant mice (Fig. 1d). The expression of other families of *V1r* genes and the thickness of the layer of *V1r*-expressing VSNs are not obviously affected (Fig. 1d). The accessory olfactory bulb (AOB), which receives axonal input from VSNs, shows a normal overall size, layered organization, and rostral segregation of axons from VSNs expressing G α_{i2} (Fig. 1e, f).

Table 1 VNO-independent functions

	Genotype	
	WT	-/-
Weight		
Males (15-20 weeks) (g)	19.9 ± 3.3	19.0 ± 3.6
Females (17-19 weeks) (g)	16.0 ± 4.3	16.7 ± 4.4
Motor activity: bar hanging and balancing		
Time spent hanging on wire (s)	39.1 ± 7.6	41.9 ± 6.1
Time to right on wire (s)	2.8 ± 0.6	2.4 ± 0.5
Motor activity: screen climbing		
Time to climb to top of screen (s)	24.6 ± 2.8	26.7 ± 2.5
Locomotor activity in open field		
Total moving distance males (cm)	470.9 ± 93.2	456.8 ± 84.1
Total moving distance females (cm)	370.4 ± 85.6	422.2 ± 71.9
Depression-related behaviours		
Forced swim test (total immobility: min)	92.2 ± 14.4	84.3 ± 11.7
Tail suspension test (total immobility: min)	107.1 ± 5.6	109.2 ± 7.9
Oestrous cycling of females		
Mean number of cycles in 17 d	1.7 ± 0.1	1.8 ± 0.2
Olfactory function: latency to find cookie		
Day 1 (s)	218.3 ± 25.3	179.2 ± 25.9
Day 2 (s)	114.4 ± 20.6	89.9 ± 17.4
Day 3 (s)	86.0 ± 9.5	87.0 ± 10.3
Day 4 (s)	56.4 ± 6.4	58.3 ± 9.3

Body weight of males and females ($n = 30$ per genotype and sex), motor activity ($n = 20$ per genotype), locomotor activity ($n = 20$ per genotype and sex), depression-related behaviours ($n = 15$ per genotype and per test) and the number of oestrous cycles of non-group-housed females ($n = 15$ per genotype) were not significantly different between wild-type (WT) and mutant (-/-) mice as evaluated with an unpaired Student's *t*-test. The test of olfactory function was repeated over four consecutive days. Both genotypes learn to find the cookie faster with repeated trials. No significant differences were observed between the genotypes ($n = 15$ per genotype and sex), as evaluated with a two-way ANOVA for repeated measurements for main effects of genotype and test day and their interaction.

‡ Present address: Department of Zoology and Animal Biology, Faculty of Sciences, University of Geneva, CH-1211 Geneva 4, Switzerland.

The $\Delta V1rab\Delta$ strain was established and analysed in an inbred 129/SvEv background. The mutant mice are viable, fertile and overtly normal. Gross body weight throughout development, motor activity, open-field locomotor activity, depression-related behaviours, and oestrous cycling of females are indistinguishable between mutant and wild-type mice (Table 1). Mutant mice do not show deficits in a food-finding test, a global measure of function of the main olfactory system (Table 1). Thus, $\Delta V1rab\Delta$ mice do not display dysfunctions that would complicate the interpretation of their performance in VNO-dependent behavioural tests.

Maternal aggression is strongly dependent on a functional VNO¹³. Normally, lactating females are aggressive towards intruders invading their nest area; however, all measures of aggression are significantly reduced in nursing $\Delta V1rab\Delta$ females: the latency to the first attack, the total time attacking, the number of attacks, the number of fights, and the number of tail rattles (Fig. 2a). The deficits are more pronounced immediately after exposure to the intruder (Fig. 2b) and attenuate as the test proceeds. Thus, the deleted *V1r* genes are involved in maternal aggression towards an intruder mouse, particularly in the acute onset of the behaviour. Other aspects of maternal behaviour are normal: the average time spent retrieving pups back to the nest is not significantly different between the genotypes (Fig. 2c), in accord with normal pup retrieval after VNO removal¹¹.

We assayed four behaviours in males. A first test focused on the

70-kHz ultrasound vocalizations emitted by males during the first minutes of exposure to a female mouse. Although this behaviour is attenuated when the VNO is removed^{14,15}, it is not altered in the mutant mice (Fig. 3a). Second, we assayed mutant males for their display of aggressive behaviours towards other, wild-type males. This VNO-dependent behaviour^{16,17} is not affected in the mutant mice (Fig. 3b): the percentage of aggressive mutant mice and their level of aggression (data not shown) are not significantly different from wild-type mice. Third, male–male sexual behaviour was scored during the aggression tests. It is commonly observed that socially inexperienced, wild-type mice exhibit sexual behaviour (composed of mounting and rhythmic, vigorous pelvic thrusting) towards other males. The expression of this behaviour diminishes as the mice learn to discriminate between sexes through social experience (Fig. 3c)¹⁸. Interestingly, for socially inexperienced mice (test 1) the level of male–male sexual behaviour is significantly lower in mutant compared to wild-type mice, and is comparable to that displayed by wild-type mice after social experience (tests 2 and 3). Thus, mutant males can discriminate better between sexes without prior experience, or their sexual drive is globally diminished.

A fourth assay of the mutant males tested their sexual behaviour towards females, which is also VNO-dependent^{11,17}. The percentage of males that mounted the female was lower in mutant compared to wild-type mice, in each of five consecutive tests (Fig. 3d). The effect tends to become stronger with consecutive trials. Normally the

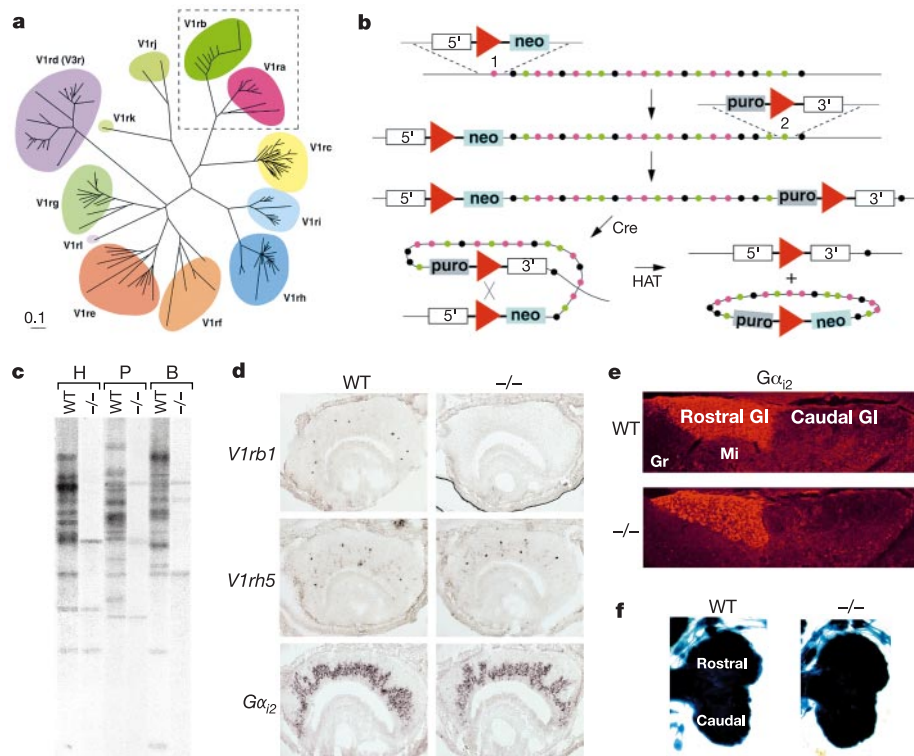


Figure 1 Targeted deletion of a cluster of *V1r* genes. **a**, The mouse *V1r* repertoire⁷. Dashed lines enclose *V1ra* and *V1rb* families. **b**, Gene targeting strategy. Top horizontal line represents the cluster^{6,12} containing *V1ra* (pink dots) and *V1rb* (green) genes. From left to right: *V1ra9*, *V1rb3*, *V1ra8*, *V1ra7*, *V1ra1*, *V1rb1*, *V1rb2*, *V1ra4*, *V1ra3*, *V1rb8*, *V1ra2*, *V1rb4*, *V1ra5*, *V1ra6*, *V1rb9*, *V1rb7*. Black dots, pseudogenes. Replacement vectors, containing the selectable marker genes *neo* or *puro*, a *loxP* site, and exons 1–2 (5') and 3–9 (3'), respectively, of the *Hprt* gene, were consecutively targeted (1 and 2) to the *V1ra9* and *V1rb7* loci in ES cells. Cre-mediated recombination between *loxP* sites and HAT selection gave cells with a reconstituted *Hprt* gene, and thus a deletion of the cluster, lost as circular DNA. The deletion encompasses all *V1ra/b* genes with an intact ORF in the

cluster, from *V1ra9* to *V1rb7* inclusive. **c**, Southern blot of genomic DNA with a *V1ra/b* probe⁸. WT, wild type; $-/-$, $\Delta V1rab\Delta$ mice; H, *HindIII*; P, *PstI*; B, *BamHI*. **d**, *In situ* hybridization with a probe of a *V1r* gene residing within the cluster (*V1rb1*, top), a *V1r* gene outside the cluster (*V1rh5*, middle), and a $G\alpha_{12}$ probe, which labels *V1r*-expressing neurons (lower). **e**, Immunostaining of an AOB sagittal section with an anti- $G\alpha_{12}$ antibody, which labels axons of *V1r*-expressing neurons projecting to the rostral region. Gl, glomerular layer; Mi, mitral cell layer; Gr, granule cell layer. **f**, Whole-mount dorsal view of the AOB of an *OMP-aulacZ* heterozygous mouse²³ (WT) and an *OMP-aulacZ* heterozygous, $\Delta V1rab\Delta$ mouse ($-/-$) stained with X-gal. In *OMP-aulacZ* mice, all VSNs and their axons are stained blue.

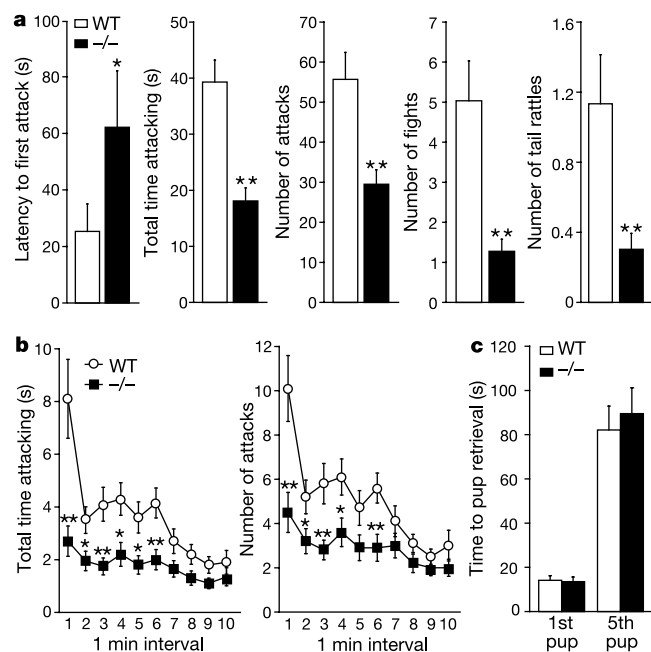


Figure 2 $\Delta V1rab\Delta$ females exhibit a reduced level of maternal aggression. **a**, Maternal aggression of WT versus mutant mice during a 10-min exposure to an unfamiliar male mouse. Data were analysed with an unpaired Student's *t*-test. **b**, Total time attacking (left) and number of attacks (right) during each 1 min interval throughout the test for WT and mutant mice. Data were analysed by a two-way analysis of variance (ANOVA) for repeated measurements for main effects of genotype and test day and their interaction, followed by *post-hoc* one-way ANOVA on each minute. Main effects of genotype, time-point, as well as interaction between these two factors were highly significant ($P < 0.001$, $P < 0.0001$ and $P < 0.002$, respectively). **c**, Performance in the pup retrieval test is not significantly different (Student's *t*-test). Data are means of 34 mice per genotype for **a**, **b** and of 35 for **c**. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$ versus WT mice.

display of sexual behaviour increases with continuous exposure to a female (Fig. 3d). This gain of sexual activity through experience does not occur in the mutant mice—instead, a decreasing tendency is seen (Fig. 3d). The diminution of the response with repetitive trials is in agreement with reports of sexual behaviour in males upon surgical removal of the VNO¹⁷ and is consistent with the proposed view of the VNO as intrinsically rewarding^{1,11}. The percentage of mutant mice that ejaculated was half of that of wild-type mice. However, if a mutant male mounted a female during a test, it displayed a level of sexual activity comparable to that of wild type in terms of the latency, number of mounts, and whether or not it ejaculated (data not shown). Thus, although fertile, mutant males have a reduced drive to initiate sexual encounters. Wild-type and mutant mice showed no differences in the levels of plasma testosterone after exposure to female mice (10.5 ± 1.2 versus 9.3 ± 1.6 ng ml⁻¹, $n = 12$ and $n = 11$ wild-type and mutant mice, respectively), ruling out a deficient testosterone response as an explanation.

The behavioural dysfunction of $\Delta V1rab\Delta$ mice may result from peripheral chemosensory deficits caused by the absence of a subset of V1r receptors. To test this hypothesis, we used an intact VNO preparation to record local field potentials (the electro-vomeranosogram, EVG) from the surface of the sensory epithelium⁹. VSNs in the apical layer of mouse VNO exhibit narrow tuning profiles for stimulus discrimination⁹. If V1r receptors function as chemoreceptors in these neurons, deletion of *V1r* genes should cause discrete deficits in the ability of the VNO to detect specific molecules. We compared field potentials in wild-type and mutant mice in response to eight pheromonal ligands known to activate apical VSNs (ref. 9,

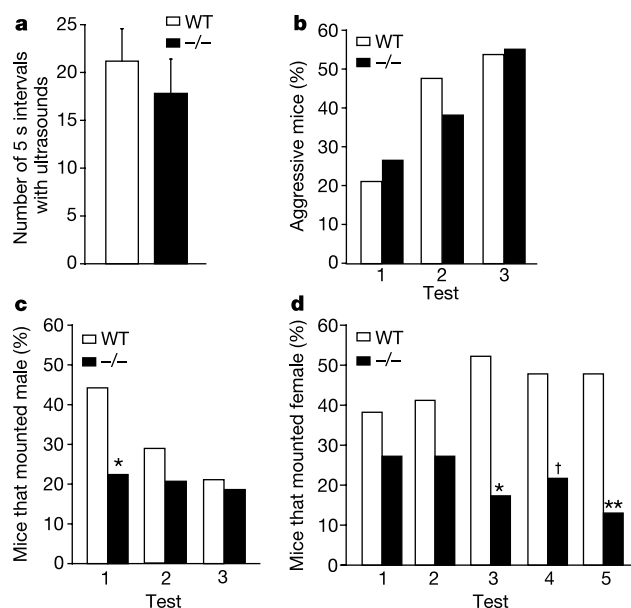


Figure 3 Behaviour of $\Delta V1rab\Delta$ males. **a**, Ultrasonic vocalizations. Mean number of 5-s intervals containing male ultrasound vocalizations to a female during a 3-min test. Data were analysed with an unpaired Student's *t*-test. $N = 20$ mice per genotype. **b**, Intermale aggression. Percentage of resident WT versus mutant mice that exhibited at least one attack towards an intruder mouse during a resident-intruder aggression test of 15 min performed in three consecutive days. $N = 45$ mice per genotype. **c**, Male-male sexual behaviour. Percentage of male mice that mounted another male at least once during the same tests performed in **b**. $N = 45$ mice, same as in **b**. **d**, Male-female sexual behaviour. Percentage of mice that mounted a receptive female at least once during each of 5 tests of 30 min duration. $N = 35$ WT and 34 mutant mice. Differences in percentages were tested with χ^2 . Dagger, $P < 0.07$; asterisk, $P < 0.05$, double asterisk, $P < 0.01$ versus WT mice.

and T.L.-Z. and F.Z., data not shown) (Fig. 4). VSNs from $\Delta V1rab\Delta$ mice do not respond to 6-hydroxy-6-methyl-3-heptanone, *n*-pentyl acetate and isobutylamine, but are otherwise normal in their responses to the other tested compounds (Fig. 4). Thus, as measured with this technique, deletion of the *V1r* gene cluster abolishes the electrophysiological response to three out of eight ligands tested. We introduce the term 'specific anosmia' for these specific chemosensory deficits, by analogy with 'specific anosmia' in the main olfactory system¹⁹. The small proportion of VSNs (<1%) in the apical layer of wild-type mice that respond physiologically to a given pheromone with a calcium increase⁹ suggests that a single *V1r* gene may correspond to each specific anosmia.

Our analyses of $\Delta V1rab\Delta$ mice provide two complementary lines of evidence for a pheromone receptor function of V1r receptors. First, mutant males have a reduced sexual drive and mutant females display a reduced level of maternal aggression. Second, their VNO exhibits specific chemosensory deficits. A chemosensory role of V1r receptors is consistent with immunolocalization of V1r receptors to microvilli of VSNs²⁰, and can be proved conclusively by functional gene transfer; we have reduced the complexity of this task to three ligands and 16 receptors. An indirect, but not mutually exclusive role is based on the observation that deletion of the *V1r* ORF precludes axonal convergence to glomeruli and results in fewer VSNs expressing the gene^{21,22}. This effect is also seen with odorant receptors, which have dual roles both in chemosensation and axon guidance²³. The impairments that result from the deletion of only ~12% of the functional *V1r* repertoire (or ~5% of the combined *V1r* and *V2r* repertoires) attest to a low level of functional redundancy within the *V1r* repertoire, in contrast to common notions

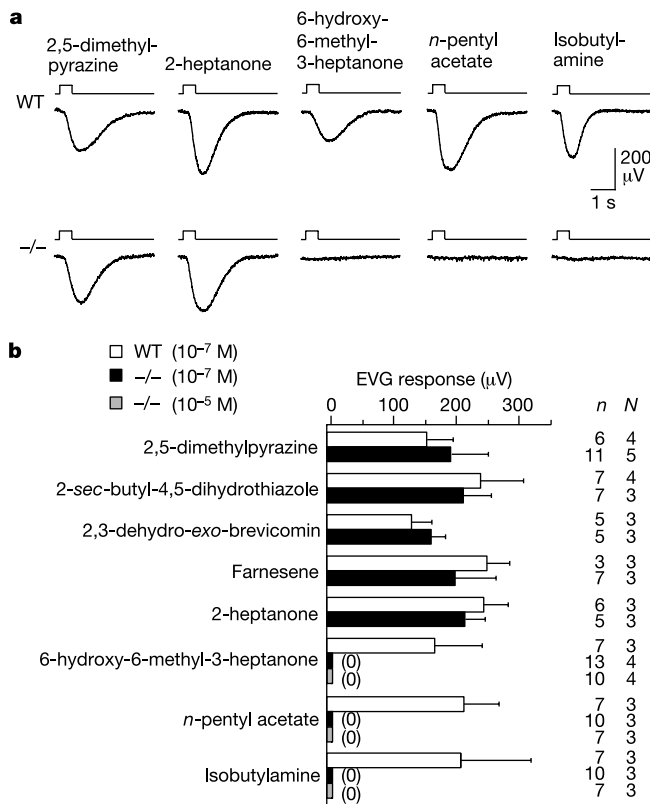


Figure 4 Chemosensory deficits in $\Delta V1rab\Delta$ mice. **a**, Comparison of EVG responses from WT (upper traces) and mutant mice (lower traces). Field potentials were induced by 500-ms pulses of the ligands shown (all at 10^{-7} M). **b**, Histograms showing collected results (mean $\pm$ s.d.), based on recordings from 9 WT mice (5 females, 4 males) and 11 mutant mice (5 females, 6 males). Peak EVG responses from both sexes were pooled. Both male and female mutant mice were defective in sensing 6-hydroxy-6-methyl-3-heptanone, *n*-pentyl acetate, and isobutylamine ($P < 0.0001$). There is no significant difference between WT and mutant mice in the detection of the other five tested ligands ($P = 0.1$ – 0.33). Stimulus concentration, 10^{-7} M (white and black bars) or 10^{-5} M (grey bars). *n*, number of recordings; *N*, number of mice.

about odorant receptors and the main olfactory system^{24,25}. Functional specialization is consistent with the phylogenetic isolation of the *V1r* families⁷, and with the narrow tuning of VSNs located in the apical layer to single ligands⁹. As pheromones in other species often consist of blends of compounds that must be present in carefully balanced ratios to elicit distinct behaviours^{26,27}, the deletion of *V1ra* and *V1rb* gene families may produce specific anosmias to certain critical components, leading to an altered representation of the blend in the brain and affecting behaviour. Deletion of a *V1r* subset may 'corrupt' pheromone coding instead of blocking it, and may thus result in phenotypes that are qualitatively distinct from those produced by VNO removal, a total receptor deletion or a global deficiency in cell signalling, such as reported in *Trp2* knockout mice^{28,29}. It will be interesting to determine if deletion of other groups of *V1r* or *V2r* genes produces distinct behavioural and electrophysiological phenotypes. Single *Vr* gene knockouts may suffice to cause deficits. □

Methods

Genetic manipulation

Replacement vectors for each end of the *V1r* gene cluster were constructed using the chromosome engineering cassettes *Hprt* $\Delta 5'$ and *Hprt* $\Delta 3'$ (ref. 5). The vectors were electroporated consecutively into AB2.2 embryonic stem cells³⁰, derived from an *Hprt*-

deficient 129/SvEv blastocyst. Clones with the double mutation were electroporated with the Cre-expression plasmid pOG231. The vectors were constructed such that recombination between the *loxP* sites reconstitutes a functional *Hprt* gene. Cells that had undergone recombination and thus expressed *Hprt* were selected in hypoxanthine-aminopterin-thymidine (HAT) medium. Sib selection analysis showed that HAT-resistant clones had regained sensitivity to neomycin and puromycin, indicating that a recombination *in cis* had occurred and confirming the deletion event. Male chimaeras were crossed to 129/SvEv females (Taconic) to maintain an inbred genetic background more suitable and reproducible for the behavioural experiments. As controls, we used wild-type offspring from heterozygous intercrosses, or wild-type offspring from wild-type parents specifically bred for this purpose. Offspring lacking the mutation were bred in parallel to produce wild-type mice. The *V1ra/b* probe used for Southern blots consisted of a pool of PCR products obtained from BACs covering the cluster by using degenerate primers corresponding to conserved regions of *V1ra/b* genes, as described⁶. *In situ* hybridizations were done with digoxigenin-labelled probes as described previously²².

Phylogenetic tree

Was constructed as in ref. 7, except that *V1rb11* was removed from the tree, because this sequence is not longer present in the updated Celera database.

Mice

Mutant and wild-type mice were in a 129/SvEv inbred background. They were housed in micro-isolator cages and given food and water *ad libitum*. Males and females were kept in separate cubicles after weaning. Behavioural experiments were performed during the dark phase of a 12:12 h light/dark cycle starting 2 h after the lights went off. Behavioural test sessions were recorded on videotape and subsequently analysed offline by an author who was unaware of the genotype of the mice. Males were weaned at 21 d old, and housed individually until behavioural testing, which was at 4–7 months. Females used for maternal aggression experiments were weaned at 21 d and housed in groups of five until they were mated, which was at 3–4 months.

Olfactory function

Mice were food-deprived for 18 h, after which an Oreo cookie was buried in the home cage of the subject under ~1 cm of bedding. The location of the cookie systematically varied across days.

Maternal aggression

Females were housed with 129/SvEv wild-type males for 15 d, after which the male was removed and females were singly housed until the day of the test. At parturition, pups were culled to 5 per litter. A single test for aggression was carried out 6–8 d after delivery. Each female was exposed for 10 min to an unfamiliar, group-housed, Swiss Webster mouse that had undergone olfactory bulbectomy (SWOBX). We recorded: (1) latency to the first attack; (2) total time attacking; cumulative time spent in any form of aggression (very rapid attacks that lasted less than 1 s were counted as 0.5 s); (3) number of attacks; number of times that the female bit the male; (4) number of fights; aggressive attacks involving intense body contact, with more than 3 consecutive seconds of tumbling, biting and wrestling; and (5) number of tail rattles.

Pup retrieval

Pups, 3 d old, were removed from their lactating mother for 5 min and returned to the home cage in a randomly distributed arrangement.

Ultrasound vocalizations

These were monitored with a bat detector (Pettersson D200). A 3-min test began by placing a C57BL/6J wild-type female into the home cage of the subject.

Intermale behaviour

Aggressive behaviours were tested three times in a resident-intruder paradigm¹⁸. Each male was tested in his home cage against a group-housed SWOBX male intruder mouse for 15 min. The latency to the first attack, the number of attacks, the cumulative duration of aggression, the number of tail rattles, and the sexual behaviour by the resident mice towards intruder mice was scored. An attack was scored each time the resident male bit the intruder male. Whenever sustained attacks occurred with repeated biting, a new attack was not scored until the previous one has ceased for at least 2 s. The cumulative duration of aggression was scored as the total time that the resident mice spent in any of the following attacks: biting, chasing, wrestling, boxing and tail rattling. For sexual behaviour, the number of times the resident mounted the intruder was scored. The same experiments were repeated on a smaller scale using a castrated male as an intruder that had been painted with urine from gonadally intact males, obtaining similar results.

Male sexual behaviour

Male mice were tested 5 times (with an interval of 1 week between each test) for 30 min with a Swiss Webster female mouse in the male's home cage. Ejaculation was measured in the last behavioural session, over an extended period of 90 min. Females were ovariectomized and subcutaneously injected with 1 μ g of oestradiol benzoate (48 h before the tests) and 500 μ g progesterone (4–7 h before the tests) to ensure high sexual receptivity. The number of attempted mounts, the latency to mount, and the number of mounts with intromissions and ejaculations were recorded. Testosterone levels were measured with standard radioimmunoassay techniques.

Electrophysiology

Mice were used at 4–8 weeks old. Local field potentials (EVG) were recorded from the luminal surface of intact VNO sensory epithelia⁹. 2,5-dimethylpyrazine, 2-heptanone, *n*-pentyl acetate and isobutylamine were purchased from Aldrich; 2-sec-butyl-4,5-dihydrothiazole from Alfa Aesar; and farnesene from Bedoukian Research.

Received 24 January; accepted 20 June 2002; doi:10.1038/nature00955.

- Halpern, M. The organisation and function of the vomeronasal system. *Annu. Rev. Neurosci.* **10**, 325–362 (1987).
- Keverne, E. B. The vomeronasal organ. *Science* **286**, 716–720 (1999).
- Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).
- Tirindelli, R., Mucignat-Caretta, C. & Ryba, N. J. Molecular aspects of pheromonal communication via the vomeronasal organ of mammals. *Trends Neurosci.* **21**, 482–486 (1998).
- Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720–724 (1995).
- Del Punta, K., Rothman, A., Rodriguez, I. & Mombaerts, P. Sequence diversity and genomic organization of vomeronasal receptor genes in the mouse. *Genome Res.* **10**, 1958–1967 (2000).
- Rodriguez, I., Del Punta, K., Rothman, A., Ishii, T. & Mombaerts, P. Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nature Neurosci.* **5**, 134–139 (2002).
- Holy, T. E., Dulac, C. & Meister, M. Responses of vomeronasal neurons to natural stimuli. *Science* **289**, 1569–1572 (2000).
- Leinders-Zufall, T. *et al.* Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792–796 (2000).
- Sam, M. *et al.* Odorants may arouse instinctive behaviours. *Nature* **412**, 142 (2001).
- Wysocki, C. J. & Lepri, J. J. Consequences of removing the vomeronasal organ. *J. Steroid Biochem. Mol. Biol.* **4**, 661–669 (1991).
- Lane, R. P., Cutforth, T., Axel, R., Hood, L. & Trask, B. J. Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion. *Proc. Natl Acad. Sci. USA* **99**, 291–296 (2002).
- Bean, N. J. & Wysocki, C. J. Vomeronasal organ removal and female mouse aggression: the role of experience. *Physiol. Behav.* **45**, 875–882 (1989).
- Bean, N. J. Olfactory and vomeronasal mediation of ultrasonic vocalizations in male mice. *Physiol. Behav.* **28**, 31–37 (1982).
- Wysocki, C. J., Nyby, J., Whitney, G., Beauchamp, G. K. & Katz, Y. The vomeronasal organ: primary role in mouse chemosensory gender recognition. *Physiol. Behav.* **29**, 315–327 (1982).
- Bean, N. J. Modulation of agonistic behavior by the dual olfactory system in male mice. *Physiol. Behav.* **29**, 433–437 (1982).
- Clancy, A. N., Coquelin, A., Macrides, F., Gorski, R. A. & Noble, E. P. Sexual behavior and aggression in male mice: involvement of the vomeronasal system. *J. Neurosci.* **4**, 2222–2229 (1984).
- Ogawa, S. *et al.* Abolition of male sexual behaviors in mice lacking estrogen receptors α and β ($\alpha\beta$ ERKO). *Proc. Natl Acad. Sci. USA* **97**, 14737–14741 (2000).
- Amoore, J. E. & Steinle, S. In *Chemical Senses* Vol. 3, *Genetics of Perception and Communication* (eds Wysocki, C. J. & Klare, M. R.) 331–351 (Marcel Dekker, New York, 1991).
- Takigami, S. *et al.* The expressed localisation of rat putative pheromone receptors. *Neurosci. Lett.* **272**, 115–118 (1999).
- Belluscio, L., Koentges, G., Axel, R. & Dulac, C. A map of pheromone receptor activation in the mammalian brain. *Cell* **97**, 209–220 (1999).
- Rodriguez, I., Feinstein, P. & Mombaerts, P. Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. *Cell* **97**, 199–208 (1999).
- Mombaerts, P. Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286**, 707–711 (1999).
- Firestein, S. How the olfactory system makes sense of scents. *Nature* **413**, 211–218 (2001).
- Slotnick, B. & Bodyak, N. Odor discrimination and odor quality perception in rats with disruption of connections between the olfactory epithelium and olfactory bulbs. *J. Neurosci.* **22**, 4205–4216 (2002).
- Hildebrand, J. G. Analysis of chemical signals by nervous systems. *Proc. Natl Acad. Sci. USA* **92**, 67–74 (1995).
- Sorensen, P. W., Christensen, T. A. & Stacey, N. E. Discrimination of pheromonal cues in fish: emerging parallels with insects. *Curr. Opin. Neurobiol.* **8**, 458–467 (1998).
- Leybold, B. G. *et al.* Altered sexual and social behaviors in trp2 mutant mice. *Proc. Natl Acad. Sci. USA* **99**, 6376–6381 (2002).
- Stowers, L., Holy, T. E., Meister, M., Dulac, C. & Koentges, G. Loss of sex discrimination and male aggression in mice deficient for TRP2. *Science* **295**, 1493–1500 (2002).
- Matzuk, M. M., Finegold, M. J., Su, J. G., Hsueh, A. J. & Bradley, A. Alpha-inhibin is a tumour-suppressor gene with gonadal specificity in mice. *Nature* **360**, 313–319 (1992).

Acknowledgements

We thank R. Peraza and A. Walsh for producing chimaeric mice, D. Pfaff for fostering this collaboration, M. Novotny for providing 2,3-dehydro-*exo*-brevicomine and 6-hydroxy-6-methyl-3-heptanone, and T. Bozza, S. Firestein, C. Greer, K. Kelliher and D. Pfaff for critical reading of the manuscript. Postdoctoral fellowship support to I.R. was from the Swiss National Foundation for Research. Grant support to T.L.-Z. and F.Z. was from NIH/NIDCD, and to P.M. from the March of Dimes Birth Defects Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.M. (e-mail: peter@rockefeller.edu).

The *ELF4* gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*

Mark R. Doyle^{*,†}, Seth J. Davis^{†,‡}, Ruth M. Bastow^{‡,§},
Harriet G. McWatters[‡], László Kozma-Bognár^{||}, Ferenc Nagy^{||},
Andrew J. Millar[‡] & Richard M. Amasino^{*}

^{*} Department of Biochemistry and Program in Cellular and Molecular Biology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

[‡] Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

^{||} Institute of Plant Biology, Biological Research Center, H-6726 Szeged, Hungary

[†] These authors contributed equally to this work

[§] Present address: John Innes Centre, Norwich NR4 7UH, UK

Many plants use day length as an environmental cue to ensure proper timing of the switch from vegetative to reproductive growth. Day-length sensing involves an interaction between the relative length of day and night, and endogenous rhythms that are controlled by the plant circadian clock¹. Thus, plants with defects in circadian regulation cannot properly regulate the timing of the floral transition². Here we describe the gene *EARLY FLOWERING 4* (*ELF4*), which is involved in photoperiod perception and circadian regulation. *ELF4* promotes clock accuracy and is required for sustained rhythms in the absence of daily light/dark cycles. *elf4* mutants show attenuated expression of *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), a gene that is thought to function as a central oscillator component^{3,4}. In addition, *elf4* plants transiently show output rhythms with highly variable period lengths before becoming arrhythmic. Mutations in *elf4* result in early flowering in non-inductive photoperiods, which is probably caused by elevated amounts of *CONSTANS* (*CO*), a gene that promotes floral induction⁵.

In *Arabidopsis*, a facultative long-day plant, the floral transition occurs earlier when plants are grown in long days (LD) than when they are grown in short days (SD)⁶. *elf4* mutants flower early in SD and are therefore impaired in their ability to sense day length (Table 1). In addition, *elf4* mutants have elongated hypocotyls and petioles (Table 1), particularly in SD. Both of these phenotypes may result from defects in circadian regulation^{7,8}. A hallmark of circadian regulation is the persistence of robust, accurate rhythms for many days under conditions of continuous light (LL) or continuous darkness (DD). Circadian outputs, including rhythmic leaf movements and the expression of chlorophyll *a/b*-binding protein (*CAB*) and cold- and circadian-regulated (*CCR*) genes, are markers of clock function.

We introduced the luciferase reporter gene fusions *CAB-LUC* and *CCR2-LUC* into wild-type and *elf4* mutant plants. *CAB* rhythms in wild-type persisted in LL, whereas *elf4* mutants lost rhythmicity after one 24-h cycle (Fig. 1a). The arrhythmia of the

Table 1 Flowering time and hypocotyl lengths of *elf4* mutants

	Total leaf number*		Hypocotyl length (mm)†		
	LD	SD	CL	LD	SD
Wild type	7.8 ± 0.4 (30)	30.6 ± 5.1 (28)	2.9 ± 0.7	2.2 ± 0.7	4.7 ± 1.3
<i>elf4</i>	6.4 ± 0.9 (36)	10.1 ± 1.1 (23)	2.4 ± 0.4	3.1 ± 0.7	9.0 ± 1.9
<i>elf3-4</i>	4.9 ± 0.3 (36)	5.9 ± 0.3 (23)	2.8 ± 0.6	6.0 ± 0.2	10.2 ± 2.2

* Mean ± s.d. total leaf number was counted as rosette leaves plus cauline leaves. The number of plants is given in parentheses.

† About 15 plants were analysed in each trial.

CL, continuous light; LD, 16 h light/8 h dark; SD, 8 h light/16 h dark.

population (Fig. 1a) concealed weak rhythms of variable period in individual *elf4* seedlings: many *elf4* seedlings had shorter or longer periods than those of wild-type seedlings (Fig. 1b). Unlike *CAB-LUC*, *CCR2-LUC* maintains a robust pattern of expression in DD. *elf4* seedlings showed weak *CCR2* expression rhythms of variable period in DD (Fig. 1d), which, like *CAB*, seemed arrhythmic in the population mean (Fig. 1c). Expression of *CCR2* also had variable periods in individual seedlings in LL (data not shown). For each marker, the rhythms of individual *elf4* seedlings damped before the rhythms of wild-type seedlings (data not shown).

Leaves of *Arabidopsis* show circadian movements that result in a

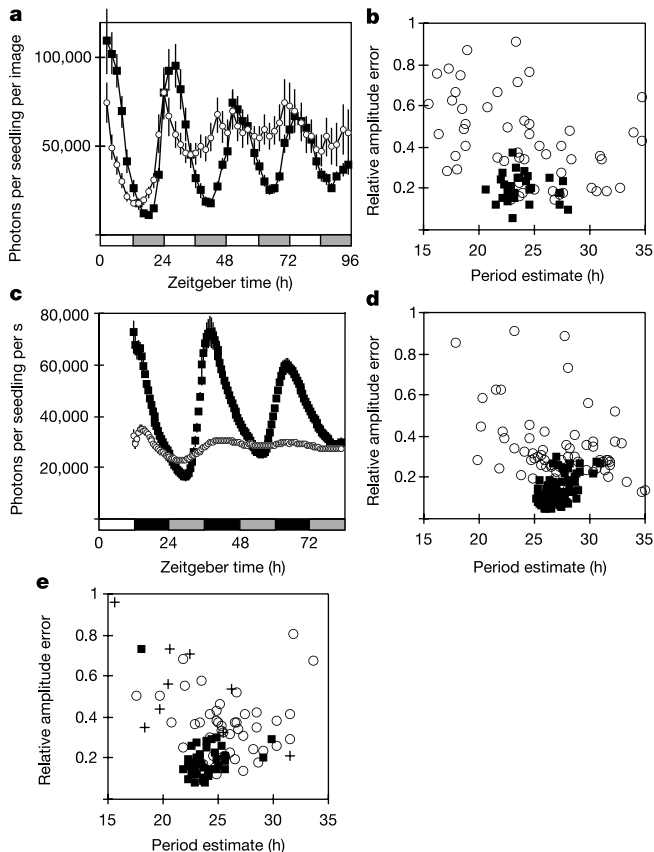


Figure 1 Circadian rhythms of *elf4* in LL or DD. **a**, *CAB-LUC* activity in plants transferred to LL at 0 h (predicted dawn). The transgene was crossed into *elf4*. Data are the means $\pm$ s.e.m. of eight individual seedling traces. **b**, *CAB-LUC* activity in LL in individual seedlings. The period estimate for each seedling is plotted against the relative amplitude error, a measure of rhythmic robustness that varies from 0 (a perfect cosine wave) to 1 (not statistically significant). Wild-type seedlings (30/30) were strongly rhythmic with a period of 23.8 ± 1.6 h (mean $\pm$ s.d.), whereas *elf4* seedlings (55/71) were weakly rhythmic with a period of 22.3 ± 5.1 h. **c**, *CCR2-LUC* activity in plants transferred to DD at 12 h (predicted dusk). At least 40 plants per line were tested for three independently transformed wild-type lines and four *elf4* lines. Data are normalized to the genotype mean; error bars show the standard deviation of each line from the genotype mean. **d**, Analysis of *CCR2-LUC* activity in DD. Wild-type (87/87) and *elf4* seedlings (67/89) were rhythmic, with periods of 26.4 ± 0.8 h and 27.1 ± 4.1 h, respectively. **e**, Rhythmic leaf movement in LL. Wild-type (45/45) and *elf4* (50/60) traces were rhythmic, with periods of 25.1 ± 1.3 h and 25.9 ± 3.2 h, respectively. In *elf3*, only 9 out of 20 traces were rhythmic, with a period of 22.5 ± 4.8 h. Filled squares, wild type; open circles, *elf4*; plus signs, *elf3*. **a** and **c** are representative of two independent experiments; **b**, **d**, **e** show pooled data representing three, two and three experiments, respectively. Grey bars represent subjective day and subjective night in **a** and **c**, respectively.

horizontal position during the day and a more vertical position at night. A large percentage of *elf4* seedlings showed rhythmic leaf movements for the first 1–2 d in LL; however, these rhythms were less robust than in wild type and were no longer detectable after 3 d (Supplementary Information Fig. 1). Similar to gene expression studies, the period of leaf movement in *elf4* varied greatly among individuals (Fig. 1e). *ELF4* is therefore important for the maintenance and accuracy of circadian rhythms under LL and DD, as tested through different outputs.

Early flowering, hyperelongation and impaired circadian rhythms in LL are all characteristics of *elf3* mutants^{9,10}. In wild-type plants, *ELF3* sustains rhythmicity in long photoperiods and in LL by inhibiting phototransduction at dusk^{11–13}. In contrast to *elf4*, *CAB-LUC* and *CCR2-LUC* expression in *elf3* mutants resembles that of the wild type in DD^{13,14}. Also, *CAB-LUC* expression and leaf movement become arrhythmic immediately under LL in *elf3* mutants^{8,15}, rather than after one cycle as in *elf4* mutants (Fig. 1a). Thus, the function of *ELF4* is distinct from that of *ELF3*.

Although the *elf4* phenotype is unique in plants, the phenotype is similar to *frequency* (*frq*) mutants of *Neurospora crassa*. *Frq* participates in a rhythmic, transcriptional feedback loop that is essential for normal circadian rhythmicity¹⁶. Mutants that are null for *frq* can seem arrhythmic in constant conditions, but weak conidiation rhythms have been detected in these mutants, with periods that are distributed broadly about the wild-type mean¹⁷. This resembles the phenotype that we have shown for *elf4*. It is unclear whether *Frq* principally generates rhythms or organizes *Frq*-independent oscillators¹⁶. In an analogous fashion, *ELF4* could function in either process in the plant clock.

Arabidopsis genes that are thought to function in the circadian oscillator include *TIMING OF CAB EXPRESSION 1* (*TOC1*), which is expressed early in the night¹⁸, and *CCA1*, which is expressed

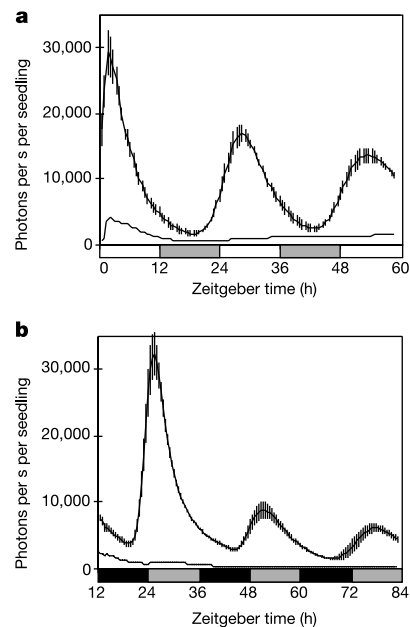


Figure 2 Regulation of *CCA1* expression by *ELF4*. Wild-type (line with error bars) and *elf4* (continuous line) plants containing a *CCA1-LUC* fusion assayed in LL (**a**) or DD (**b**). At least 18 plants per line in **a** or at least 40 plants per line in **b** were tested for five independently transformed wild-type and six *elf4* lines. *CCA1* expression was low in *elf4* (in LL: wild-type, $8,938 \pm 1,631$ counts per seedling per s (mean $\pm$ s.e.m.); *elf4*, $1,330 \pm 478$; in DD: wild-type $7,178 \pm 927$; *elf4*, 613 ± 156). Data are normalized to the genotype mean and are representative of two independent experiments.

shortly after dawn⁴. The *elf4* lesion disrupted rhythmic *CCA1* expression, showing the importance of *ELF4* in circadian regulation. The expression of a *CCA1-LUC* reporter gene was low and arrhythmic in *elf4* seedlings under both LL and DD (Fig. 2). *CCA1* expression in *elf4* occurred only through acute light activation after the dark/light transition (Fig. 2a). This low amplitude pulse of *CCA1* might have supported the single cycle of *CAB-LUC* and *CCR2-LUC* expression in *elf4* in LL (Fig. 1a and data not shown), before the normal rhythm was lost around subjective dawn of the subsequent cycle, when *CCA1* was again required but was not expressed. We propose that the circadian function of *ELF4* is to activate light-inducible clock components around dawn: *CCA1* could be one such component. *TOC1* may carry out a similar activation function³.

elf4 was derived from a population mutagenized with the transfer DNA (tDNA) of *Agrobacterium tumefaciens*¹⁹. Adjacent to the tDNA was a deleted region predicted to contain a small open reading frame (ORF). A clone containing this ORF was sufficient to restore the phenotypes of *elf4* to that of wild type (Supplementary Information Fig. 2). In LL, the *ELF4* transcript maintains a robust rhythm, which shows that transcript abundance is under circadian control (Fig. 3a,

b). Robust cycling is not detectable after transfer to DD (Fig. 3a, c), which is similar to what has been observed for *TOC1* and *ELF3* (refs 12, 18). *ELF4* also shows a photoperiod-specific pattern of expression. In LD, messenger RNA levels peak around 12 h after dawn (Fig. 3d, e). In SD, the peak shifts to around 8 h after dawn (Fig. 3d, e). The predicted *ELF4* protein is 111 amino acids and does not contain known protein signatures (Supplementary Information Fig. 3). So far, organisms others than plants have not been found to contain genes similar to *ELF4*.

The circadian clock is thought to function in the floral transition through a coincidence mechanism in which inductive stimuli must be present during a particular phase of the rhythm to be effective¹. Given its strong effect on circadian outputs, *ELF4* could repress flowering by controlling the phase of an output rhythm specific for flowering-time regulation. The floral-promoting gene *CO* may act in such an output^{5,20,21}. The diurnal pattern of *CO* mRNA accumulation is distinct in inductive versus non-inductive photoperiods, and this expression pattern could contribute to a coincidence mechanism²¹. Several circadian rhythm mutants affect *CO* expression in ways that correlate with their early or late flowering phenotypes²¹. In SD-grown *elf4* mutants, the level of *CO* mRNA is increased at all time points (Fig. 3f, g). Increased *CO* expression is consistent with the early flowering *elf4* phenotype²¹.

The molecular circuitry that comprises the *Arabidopsis* circadian clock is just beginning to be understood³. Our gene expression and leaf movement assays have revealed unique phenotypes of *elf4*, which indicate that *ELF4* functions either in a circadian oscillator or by conferring accuracy and persistence on a separate oscillator. Despite a unique role in clock regulation, the effect of *ELF4* on flowering time may occur by a mechanism similar to that of other circadian mutants, that is, through the misexpression of *CO*²¹. Further characterization of the *elf4* mutant will hopefully elucidate how components of the circadian machinery collectively control rhythmic processes throughout the day/night cycle, including the ability of plants to respond to photoperiod. □

Methods

Plant material and growth and conditions

All experiments were carried out on the Ws ecotype. Plants were grown in LD (16 h light/8 h dark) or SD (8 h light/16 h dark) under cool white light ($\sim 120 \mu\text{mol m}^{-2} \text{s}^{-1}$) at a constant 22 °C unless indicated otherwise. Hypocotyl-growth assays were done as described²².

Analysis of *ELF4* and *CO* mRNA abundance

We isolated RNA from 1-week-old whole seedlings using TRI reagent (Sigma). Complementary DNA was synthesized by polymerase chain reaction with reverse transcription (RT-PCR) using 5 μg of total RNA, an oligo dT primer with an M13 sequence attached to the 5' end and Superscript II reverse transcriptase (Gibco). As *ELF4* is a single exon gene, RT-PCR was carried out using a gene-specific primer, 5'-TATGAAGA GGAACGGCGAGA-3', and the M13F primer, 5'-GTAAACGACGCGCCAGTCCC-3', to prevent DNA contamination. *UBQ10* (ref. 23) was amplified as a control. We used 25 and 20 cycles for *ELF4* and *UBQ*, respectively. PCR fragments were transferred to a nylon membrane (Hybond) and hybridized with labelled *ELF4* and *UBQ* cDNAs. We measured *CO* mRNA abundance as described²¹. RNA blots contained 10 μg of total RNA per lane on a denaturing formaldehyde gel (1% agarose). RNA was transferred to a nylon membrane and probed with an *ELF4* antisense RNA probe.

Circadian rhythm assays

The *CAB2-LUC* construction has been described²⁴. The DNA regions immediately upstream of the *CCR2* and *CCA1* genes were amplified by PCR from *Arabidopsis* (Ws-2) genomic DNA. These promoters were fused to luciferase and introduced into *Arabidopsis*. Seeds were stratified (48 h at 4 °C) on Murashige and Skoog (MS) agar medium with 3% sucrose. Seedlings were entrained for 6 d in 12 h light ($140 \mu\text{mol m}^{-2} \text{s}^{-1}$ cool white fluorescent light)/12 h dark in controlled environment chambers (Percival or Sanyo) before being introduced into assay conditions²⁴. The luminescence of individual seedlings was measured by low-light video imaging using a liquid nitrogen cooled camera (Roper Scientific)²⁴ or by counting in a Packard Topcount²⁵. Instrument background has been subtracted from the data presented here. Periods were estimated starting from 24 h after the last dark/light transition, using FFT-NLLS software as described¹¹. We entrained seedlings for leaf movement assays for 9 d as described above. Leaf growth was monitored by video imaging under LL. Movements of the first primary leaves over the first 90 h of imaging were analysed using FFT-NLLS software as described⁶. Period means and

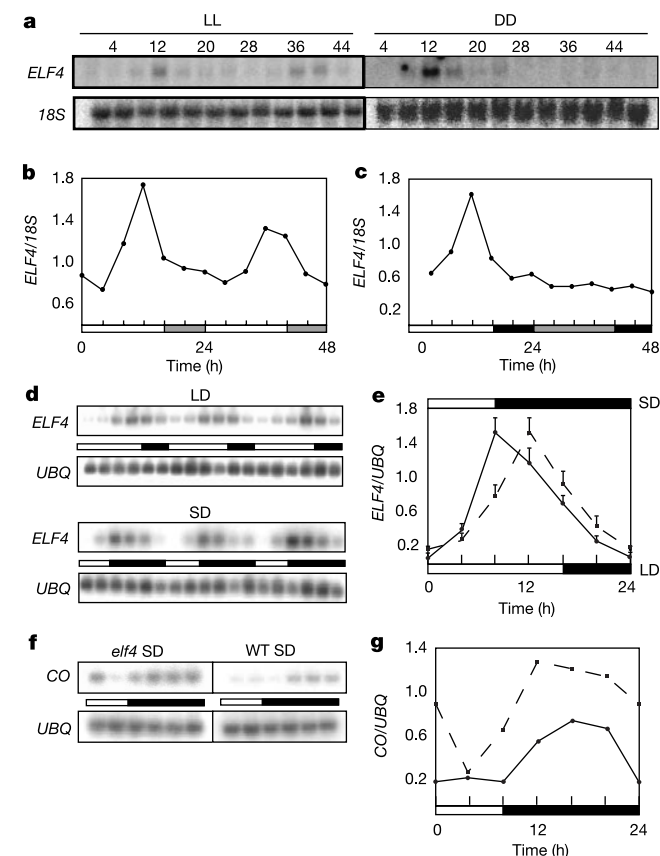


Figure 3 Expression of *ELF4* and *CONSTANS*. **a**, RNA blot of *ELF4* expression in LL and DD. Data are representative of two independent experiments. **b**, **c**, Quantification of *ELF4* expression in LL and DD. White and black rectangles indicate lights on and lights off, respectively. Grey rectangles indicate subjective day and subjective night in **b** and **c**, respectively. **d**, RT-PCR showing *ELF4* expression under LD and SD over 3 d. **e**, Trace showing *ELF4* expression in LD (dashed line) and SD (solid line). Points and error bars indicate the mean $\pm$ s.e.m. of three trials for LD and six trials for SD. **f**, RT-PCR showing *CO* expression in SD-grown wild-type (WT) and *elf4*. **g**, Trace showing *CO* expression in wild type (solid line) and *elf4* (dashed line). Data are representative of three independent experiments. The 0-h time point is plotted again at 24 h. White and black rectangles indicate lights on and lights off, respectively.

standard deviations are variance weighted. Rhythmic traces are those with any period estimate in the circadian range, 15–35 h, as described⁸. All rhythm assays were conducted at 22 °C.

Received 9 April; accepted 13 June 2002; doi:10.1038/nature00954.

1. Thomas, B. & Vince-Prue, D. *Photoperiodism in Plants* (Academic, San Diego, 1997).
2. Carré, I. A. Day-length perception and the photoperiodic regulation of flowering in *Arabidopsis*. *J. Biol. Rhythms* **16**, 416–424 (2001).
3. Alabadi, D. *et al.* Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. *Science* **293**, 880–883 (2001).
4. Wang, Z. & Tobin, E. M. Constitutive expression of the *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) gene disrupts circadian rhythms and suppresses its own expression. *Cell* **93**, 1207–1217 (1998).
5. Putterill, J., Robson, E., Lee, K., Simon, R. & Coupland, G. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**, 847–857 (1995).
6. Koornneef, M., Hanhart, C. J. & van der Veen, J. H. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **229**, 57–66 (1991).
7. Samach, A. & Coupland, G. Time measurement and the control of flowering in plants. *BioEssays* **22**, 38–47 (2000).
8. Dowson-Day, M. J. & Millar, A. J. Circadian dysfunction causes aberrant hypocotyl elongation patterns in *Arabidopsis*. *Plant J.* **17**, 63–71 (1999).
9. Hicks, K. A. *et al.* Conditional circadian dysfunction of the *Arabidopsis* early-flowering 3 mutant. *Science* **274**, 790–792 (1996).
10. Zagotta, M. T. *et al.* The *Arabidopsis* *ELF3* gene regulates vegetative photomorphogenesis and the photoperiodic induction of flowering. *Plant J.* **10**, 691–702 (1996).
11. McWatters, H. G., Bastow, R. M., Hall, A. & Millar, A. J. The *ELF3* zeitnehmer regulates light signalling to the circadian clock. *Nature* **408**, 716–720 (2000).
12. Liu, X. L., Covington, M. F., Fankhauser, C., Chory, J. & Wagner, D. R. *ELF3* encodes a circadian clock-regulated nuclear protein that functions in an *Arabidopsis* PHYB signal transduction pathway. *Plant Cell* **13**, 1293–1304 (2001).
13. Covington, M. F. *et al.* *ELF3* modulates resetting of the circadian clock in *Arabidopsis*. *Plant Cell* **13**, 1305–1315 (2001).
14. Hicks, K. A., Albertson, T. M. & Wagner, D. R. *EARLY FLOWERING 3* encodes a novel protein that regulates circadian clock function and flowering in *Arabidopsis*. *Plant Cell* **13**, 1281–1292 (2001).
15. Reed, J. W. *et al.* Independent action of *ELF3* and *phyB* to control hypocotyl elongation and flowering time. *Plant Physiol.* **122**, 1149–1160 (2000).
16. Bell-Pedersen, D., Crosthwaite, S. K., Lakin-Thomas, P. L., Merrow, M. & Okland, M. The *Neurospora* circadian clock: simple or complex? *Phil. Trans. R. Soc. Lond. B* **356**, 1697–1709 (2001).
17. Aronson, B. D., Johnson, K. A. & Dunlap, J. C. Circadian clock locus frequency: protein encoded by a single open reading frame defines period length and temperature compensation. *Proc. Natl Acad. Sci. USA* **91**, 7683–7687 (1994).
18. Strayer, C. *et al.* Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science* **289**, 768–771 (2000).
19. Feldmann, K. A. T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *Plant J.* **1**, 71–82 (1991).
20. Onouchi, H., Igeno, M. I., Perilleux, C., Graves, K. & Coupland, G. Mutagenesis of plants overexpressing *CONSTANS* demonstrates novel interactions among *Arabidopsis* flowering-time genes. *Plant Cell* **12**, 885–900 (2000).
21. Suarez-Lopez, P. *et al.* *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* **410**, 1116–1120 (2001).
22. Davis, S. J., Bhoo, S. H., Durski, A. M., Walker, J. M. & Vierstra, R. D. The heme-oxygenase family required for phytochrome chromophore biosynthesis is necessary for proper photomorphogenesis in higher plants. *Plant Physiol.* **126**, 656–669 (2001).
23. Kardailsky, I. *et al.* Activation tagging of the floral inducer *FT*. *Science* **286**, 1962–1965 (1999).
24. Bognar, L. K. *et al.* The circadian clock controls the expression pattern of the circadian input photoreceptor, phytochrome B. *Proc. Natl Acad. Sci. USA* **96**, 14652–14657 (1999).
25. Carré, I. A. & Kay, S. A. Multiple DNA–protein complexes at a circadian-regulated promoter element. *Plant Cell* **7**, 2039–2051 (1995).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

This work was supported by the College of Agricultural and Life Sciences of the University of Wisconsin and by a grant to R.A. from the National Science Foundation. M.R.D. was supported by a Molecular Biosciences Training Grant (NIH); R.M.B. was supported by a Gatsby graduate studentship; S.J.D. is a Department of Energy Bioscience fellow of the Life Sciences Research Foundation. Work in Warwick was supported by grants from the Biotechnology and Biological Sciences Research Council and the Human Frontier Science Program (HFSP) to A.J.M. The work in Hungary was supported by the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M.A. (e-mail: amasino@biochem.wisc.edu) or A.J.M. (e-mail: Andrew.Millar@warwick.ac.uk). The accession code for *ELF4* is AY035183.

Protective role of phospholipid oxidation products in endotoxin-induced tissue damage

Valery N. Bochkov, Alexandra Kadl, Joakim Huber, Florian Gruber, Bernd R. Binder & Norbert Leitinger

Department of Vascular Biology and Thrombosis Research, University of Vienna, Schwarzschanerstrasse 17, 1090 Vienna, Austria; and BMT-Research, Brunnerstrasse 59/5, 1235 Vienna, Austria

Lipopolysaccharide (LPS), an outer-membrane component of Gram-negative bacteria, interacts with LPS-binding protein and CD14, which present LPS to toll-like receptor 4 (refs 1, 2), which activates inflammatory gene expression through nuclear factor κ B (NF κ B) and mitogen-activated protein-kinase signalling^{3,4}. Antibacterial defence involves activation of neutrophils that generate reactive oxygen species capable of killing bacteria⁵; therefore host lipid peroxidation occurs, initiated by enzymes such as NADPH oxidase and myeloperoxidase⁶. Oxidized phospholipids are pro-inflammatory agonists promoting chronic inflammation in atherosclerosis⁷; however, recent data suggest that they can inhibit expression of inflammatory adhesion molecules⁸. Here we show that oxidized phospholipids inhibit LPS-induced but not tumour-necrosis factor- α -induced or interleukin-1 β -induced NF κ B-mediated upregulation of inflammatory genes, by blocking the interaction of LPS with LPS-binding protein and CD14. Moreover, in LPS-injected mice, oxidized phospholipids inhibited inflammation and protected mice from lethal endotoxin shock. Thus, in severe Gram-negative bacterial infection, endogenously formed oxidized phospholipids may function as a negative feedback to blunt innate immune responses. Furthermore, identified chemical structures capable of inhibiting the effects of endotoxins such as LPS could be used for the development of new drugs for treatment of sepsis.

Lipids containing polyunsaturated fatty acids such as 1-palmitoyl-2-arachidonoyl-*sn*-glycero-3-phosphorylcholine (PAPC) are especially prone to oxidative modification. Oxidation of PAPC results in generation of a mixture of oxidation products (OxPAPC), some of which stimulate adhesion of monocytes to endothelial cells and induce expression of inflammatory genes in endothelial cells^{8–10}. Thus, a role for oxidized phospholipids as culprits in chronic inflammation was suggested. On the other hand, we have shown that OxPAPC inhibited neutrophil adhesion to endothelial cells induced by LPS⁸.

Here we show that OxPAPC blocked LPS-induced upregulation of the inflammatory adhesion molecules E-selectin, ICAM-1 and VCAM-1 on human umbilical-vein endothelial cells (HUVEC) (Fig. 1a). The inhibitory effect of OxPAPC was concentration-dependent, with half-maximal inhibition observed at 10 μ g ml⁻¹ OxPAPC (Fig. 1b). The effects of LPS at concentrations as high as 500 ng ml⁻¹ were inhibited by 50 μ g ml⁻¹ OxPAPC. Because expression of E-selectin, ICAM-1 and VCAM-1 is NF κ B-dependent, we examined the influence of OxPAPC on signalling events at different levels of the NF κ B cascade. We found that OxPAPC inhibited LPS-induced activation of a 5 $\times$ NF κ B-luciferase reporter construct (Fig. 1c), binding of p65 to its DNA consensus site (Fig. 1d), and phosphorylation and degradation of I κ B α (Fig. 1e).

NF κ B is activated during inflammation by various receptors including the LPS toll-like receptor 4 (TLR4), the interleukin 1 (IL-1) receptor, and the tumour-necrosis factor- α (TNF α) receptor. In contrast to LPS, TNF α -induced phosphorylation and degra-

dation of I κ B α were not significantly inhibited by OxPAPC (Fig. 1e). Moreover, OxPAPC blocked LPS-induced but not TNF α -induced activation of p38 mitogen-activated protein (MAP) kinase (Fig. 1e). Signalling events downstream of the IL-1 receptor and TLR4 converge on MyD88, which binds to the cytosolic domains of the receptors and couples them to downstream signalling cascades⁴. However, effects of IL-1 β on expression of E-selectin were not inhibited by OxPAPC (Fig. 1f). Thus, these data indicate that the inhibitory effect of OxPAPC was specific for LPS and occurred upstream of MyD88, apparently at the level of LPS recognition by cellular or soluble receptor(s).

Activation of innate immune responses by LPS can be inhibited by various soluble molecules that bind LPS and neutralize its activity. These include certain plasma proteins, lipoproteins and lipids^{11–14} and inhibitors structurally related to polymyxin B¹⁵. Alternatively, the action of LPS is inhibited by lipid-A-like LPS antagonists, which form inactive complexes with TLR4 or its accessory proteins¹⁶. Both the inhibitory action of LPS-binding substances as well as lipid-A-like LPS antagonists can be overcome

by high LPS concentrations¹⁷. We found, however, that increasing LPS concentrations could not overcome the inhibitory effect of OxPAPC (Fig. 2a). Moreover, LPS-trapping substances are known to inhibit LPS-induced clotting activity of the *Limulus* amoebocyte lysate (LAL), while several lipid-A-like LPS antagonists activate clotting¹⁸. OxPAPC neither stimulated nor inhibited LPS-induced clotting activity of the LAL (Fig. 2b). Although serum also contains LPS-binding activity, increasing concentrations of serum did not mimic the inhibitory action of OxPAPC on E-selectin expression induced by LPS (Fig. 2c). These data indicate that OxPAPC inhibits LPS action by a mechanism different from known LPS antagonists. We hypothesized that OxPAPC could interfere with recognition of LPS by LPS-binding protein (LBP) and CD14, both presenting LPS to TLR4 (ref. 2)

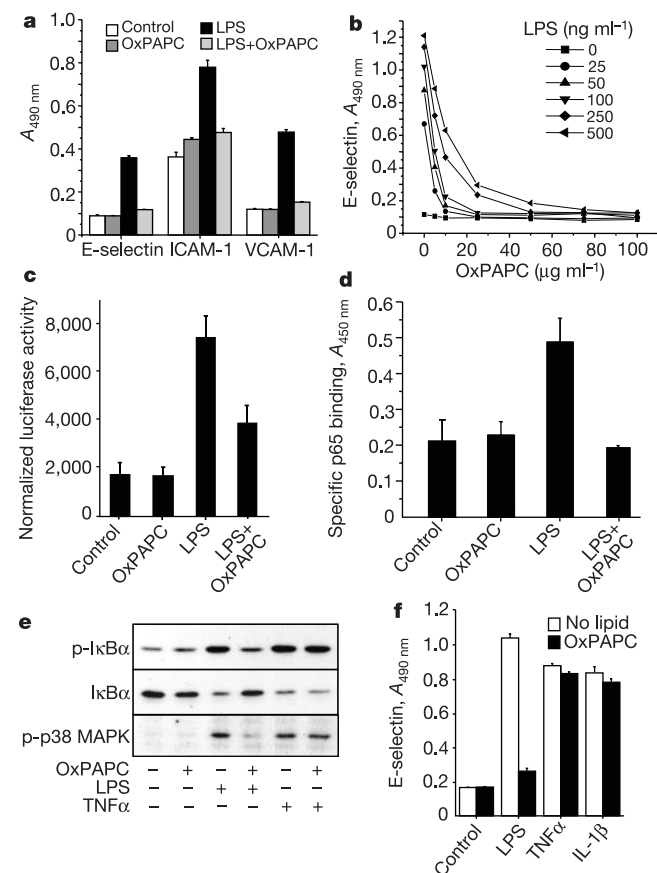


Figure 1 OxPAPC selectively inhibits LPS-induced intracellular signalling and expression of inflammatory adhesion molecules in HUVEC. **a**, OxPAPC inhibits LPS-induced upregulation of E-selectin, ICAM-1 and VCAM-1. **b**, The inhibitory effect of OxPAPC on LPS-induced E-selectin expression is concentration-dependent. **c**, OxPAPC inhibits LPS-induced activation of a $5 \times$ NF κ B-luciferase reporter. **d**, OxPAPC inhibits LPS-induced DNA-binding activity of p65. 'Specific' binding represents the difference between the binding in the absence or presence of excess of soluble consensus oligonucleotide. **e**, OxPAPC blocks LPS-induced phosphorylation (p) and degradation of I κ B α and p38 MAP kinase phosphorylation. **f**, OxPAPC inhibits the action of LPS, but not of IL-1 β and TNF α . HUVEC were stimulated with LPS (300 ng ml^{-1}), TNF α (20 U ml^{-1}) or IL-1 β (10 ng ml^{-1}) in the presence or absence of $50 \text{ } \mu\text{g ml}^{-1}$ OxPAPC. **a–f**, See Methods for additional experimental details. $A_{490 \text{ nm}}$, absorbance at 490 nm.

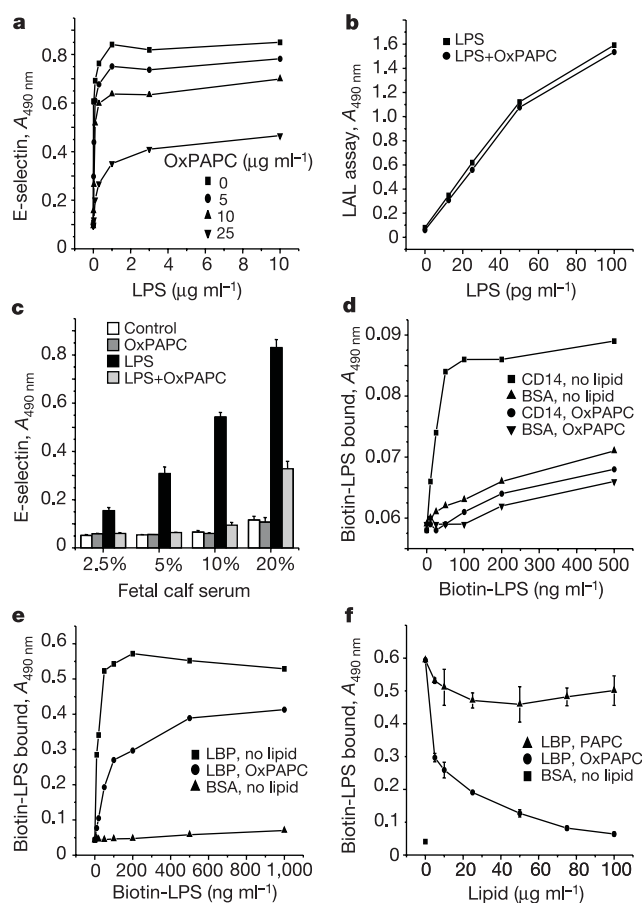


Figure 2 OxPAPC inhibits effects of LPS by blocking interactions of LPS with LBP and CD14. **a**, Increasing concentrations of LPS do not overcome the inhibitory activity of OxPAPC. HUVEC were stimulated with increasing concentrations of LPS in the presence of indicated concentrations of OxPAPC. E-selectin ELISA was performed as described in Methods. **b**, OxPAPC does not interfere with the LAL endotoxin assay. **c**, LPS-induced expression of E-selectin is not inhibited by high concentrations of serum. HUVEC were stimulated with LPS in the presence or absence of OxPAPC in medium 199 supplemented with indicated concentrations of bovine calf serum. **d**, OxPAPC inhibits binding of biotin-LPS to immobilized CD14. Binding and detection of biotinylated LPS to CD14- or BSA-coated microwell plates were performed as described in the Methods. Where indicated, the incubation medium contained $50 \text{ } \mu\text{g ml}^{-1}$ OxPAPC. **e**, OxPAPC inhibits binding of biotin-LPS to immobilized LBP. **f**, Concentration dependence of the inhibition of LPS/LBP binding by OxPAPC. Microwells containing immobilized LBP were treated with indicated concentrations of sonicated native or oxidized PAPC and 100 ng ml^{-1} biotin-LPS. Detection of bound biotin-LPS was performed as described in Methods.

and being critical for LPS signalling in endothelial cells^{19,20}. Apart from LPS, CD14 and LBP have been shown to bind phospholipids catalysing their transfer between cells and lipoproteins^{21,22}. Indeed, OxPAPC inhibited binding of LPS to immobilized CD14 (Fig. 2d). OxPAPC also inhibited binding of LPS to immobilized LBP (Fig. 2e). The inhibitory action was concentration-dependent

and specific for oxidized PAPC, whereas native PAPC was not effective (Fig. 2f).

Next we examined structural requirements for lipids to inhibit effects of LPS. We found that the inhibitory activity was specific for oxidized phospholipids because other lipids that are potentially present in OxPAPC preparations and known to exert a variety of

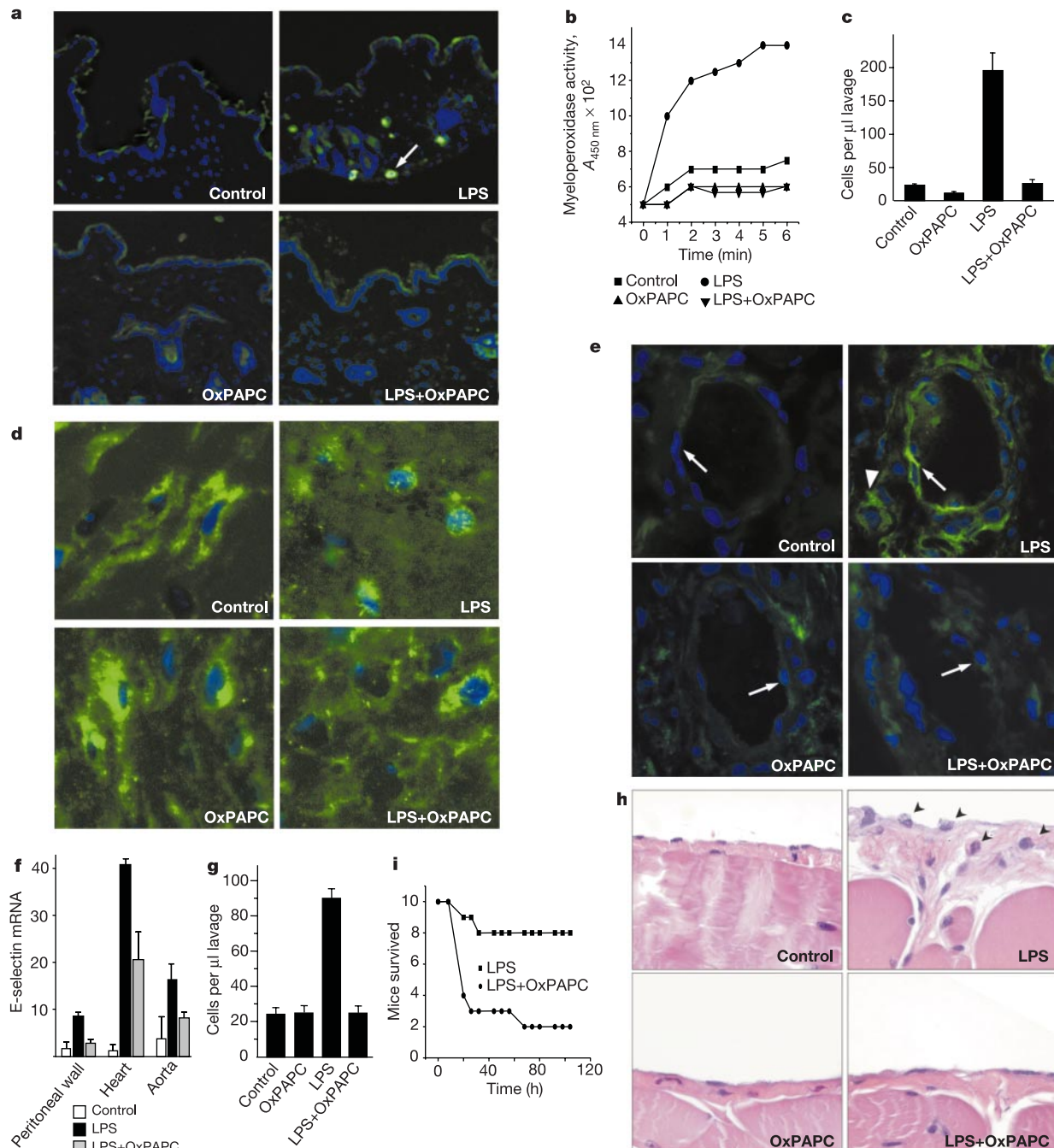


Figure 3 Oxidized phospholipids inhibit effects of endotoxin *in vivo*. **a**, OxPAPC inhibits LPS-induced accumulation of leukocytes reactive for myeloperoxidase in mouse skin. Arrow indicates myeloperoxidase-expressing leukocyte. **b**, Myeloperoxidase activity in the skin of LPS/OxPAPC-treated mice. **c**, OxPAPC inhibits LPS-induced accumulation of leukocytes in the air pouch model of inflammation. **d**, OxPAPC inhibits LPS-induced nuclear translocation of p65 in the air pouch wall. We note colocalization of p65 (green) with DAPI-stained nuclei (blue) in LPS-treated cells. **e**, OxPAPC inhibits LPS-induced VCAM-1 expression in the air pouch wall. Arrow indicates endothelial cell, arrowhead indicates myofibroblasts of the pouch wall. **f**, OxPAPC inhibits induction of E-selectin

mRNA by LPS. Expression of E-selectin mRNA in heart, peritoneal wall and aorta was measured after injection of 50 μg of LPS intraperitoneally with or without 200 μg of OxPAPC in 0.9% saline. After 3 h, animals were killed, and the tissues were processed for E-selectin mRNA quantification. **g**, OxPAPC inhibits LPS-induced extravasation of white blood cells into the peritoneum. **h**, OxPAPC inhibits LPS-induced oedema formation in the peritoneal wall. Staining is with haematoxylin–eosin. Arrowheads indicate inflammatory cells. **i**, OxPAPC increases survival in mice treated with lethal doses of LPS. **a–i**, See Methods for experimental details.

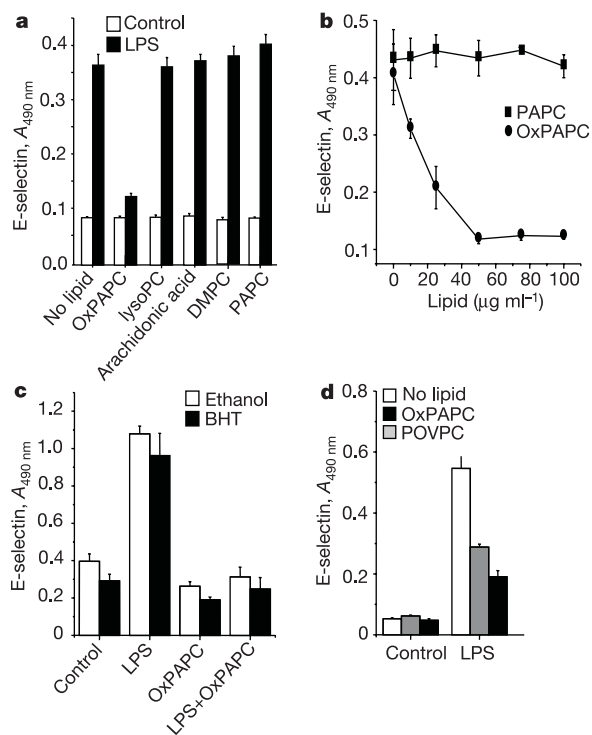


Figure 4 The anti-endotoxin activity of OxPAPC is not mimicked by other lipids, crucially depends on OxPAPC oxidation, and is not mediated by free radicals. **a**, The inhibitory effect of OxPAPC is not mimicked by other lipids. HUVEC were incubated with the indicated lipids ($30 \mu\text{g ml}^{-1}$) and 300 ng ml^{-1} LPS. **b**, Oxidation of PAPC is a prerequisite for its anti-endotoxin activity. HUVEC were treated with indicated concentrations of sonicated native or oxidized PAPC and 300 ng ml^{-1} LPS. **c**, Inhibitory action of OxPAPC is not rescued by the free-radical scavenger butylated hydroxytoluene (BHT). HUVEC were stimulated with LPS/OxPAPC in the presence of $25 \mu\text{M}$ BHT or 0.1% ethanol (vehicle). **d**, POVPC inhibits LPS-induced upregulation of E-selectin. HUVEC were treated with POVPC or OxPAPC (each at $20 \mu\text{g ml}^{-1}$) and LPS (300 ng ml^{-1}). **a–d**, After 4 h, surface expression of E-selectin was determined by cell ELISA as described in Methods.

biological effects, namely arachidonic acid and lysophosphatidylcholine, did not inhibit LPS-induced upregulation of E-selectin (Fig. 4a). Furthermore, only oxidized PAPC possessed the inhibitory activity, while non-oxidized PAPC or phosphatidylcholine containing saturated fatty acids (dimyristoylphosphatidylcholine, DMPC) were inactive (Fig. 4a and b). These data clearly show that oxidative modification is essential for phospholipids to gain anti-endotoxin activity. The inhibitory activity was not due to free-radical generation or free radicals already present in the OxPAPC preparations because addition of the antioxidant butylated hydroxytoluene had no effect (Fig. 4c). To identify individual lipid oxidation products responsible for the inhibitory activity, the lipid oxidation products contained in OxPAPC were fractionated using high-performance liquid chromatography/mass spectrometry (HPLC/MS)⁹. In fractions that inhibited LPS action, we consistently found a molecule with mass-to-charge ratio (m/z) = 594, previously identified as 1-palmitoyl-2-oxovalerolyl-*sn*-glycero-3-phosphorylcholine (POVPC)⁹. In fact, synthesized POVPC⁸ inhibited LPS-induced E-selectin expression (Fig. 4d).

To determine whether OxPAPC exerts inhibitory effects also *in vivo*, we investigated LPS-induced inflammation at various levels of the inflammatory response, that is, NF κ B activation, expression of inflammatory adhesion molecules and extravasation of inflammatory cells, using three established models of local and systemic LPS-induced inflammation in mice. In a skin model, OxPAPC inhibited LPS-induced accumulation of myeloperoxidase-expressing leuko-

cytes at the site of injection (Fig. 3a, b). In an air pouch model²³, OxPAPC inhibited LPS-induced neutrophil immigration into the pouch cavity (Fig. 3c), p65 nuclear translocation in cells of the pouch wall (Fig. 3d) and VCAM-1 expression in endothelial cells as well as myofibroblasts in the pouch wall (Fig. 3e). After intraperitoneal injection of LPS, OxPAPC inhibited induction of E-selectin messenger RNA locally in the peritoneal wall, but also systemically in heart and aorta (Fig. 3f). Moreover, OxPAPC inhibited accumulation of blood-borne cells in the peritoneal cavity (Fig. 3g) and blocked oedema formation in the subserosal layer of the peritoneal wall (Fig. 3h). These data show that OxPAPC inhibits inflammatory effects of LPS also *in vivo*. Finally, OxPAPC significantly increased survival in mice injected with lethal doses of LPS (Fig. 3i).

Taken together, these results demonstrate that inhibition of LPS-induced acute inflammation by oxidized phospholipids suppresses innate immune responses both *in vitro* and *in vivo*. Our data clearly show that the anti-inflammatory action of OxPAPC is specific for LPS as compared to inflammatory cytokines TNF α and IL-1 β . Thus, endogenously formed oxidized phospholipids potentially blunt the initiation of LPS-induced acute inflammation, although previous data indicate that they propagate chronic inflammation⁷. Furthermore, OxPAPC inhibits action of endotoxin through a mechanism apparently different from simple LPS entrapment leading to the formation of biologically inactive complexes. Our data instead suggest that the inhibition of the LPS effects by OxPAPC is mediated by interference with LPS binding to LBP and CD14, thereby inhibiting LPS presentation to TLR4.

We also find that oxidation is crucial for PAPC to acquire the anti-endotoxin activity. Thus, oxidation renders phospholipids capable of interacting with components of the innate immune system and of modifying responses to Gram-negative bacterial products. Given that under normal conditions LPS antagonistic activity of PAPC is low, it may increase dramatically at sites of bacterial inflammation owing to high concentrations of neutrophil-derived reactive oxygen species inducing lipid peroxidation⁵. This raises the possibility that phospholipid oxidation serves as a negative feedback mechanism to downregulate acute inflammation in Gram-negative infection. In addition, we identify POVPC as one chemical structure exerting anti-endotoxin effects, a compound quite different from previously described LPS scavengers or LPS receptor antagonists such as polymyxin B or lipid-A like substances^{15,16}. These findings open up the possibility of developing new anti-endotoxin drugs with distinct chemical and pharmacodynamic properties. □

Methods

Lipids and lipopolysaccharide

PAPC, DMPC, 1-palmitoyl-*sn*-glycero-3-phosphorylcholine (lysoPC), arachidonic acid and lipopolysaccharide from *E. coli* serotype 055:B5 were purchased from Sigma-Aldrich, Vienna, Austria. OxPAPC was obtained by air oxidation of dry PAPC as described previously⁹ and was stored in chloroform at -70°C . POVPC was synthesized as described⁸. Immediately before the experiment, lipids were dried in glass tubes under the stream of N_2 and solubilized in culture medium by vigorous vortexing for 30 s. Where indicated, the lipids were additionally sonicated for 30 s.

Cells

Human umbilical-vein endothelial cells (HUVEC) were obtained and cultured as described¹⁰. Monolayers of HUVEC at passages 3 to 6 were treated with OxPAPC and agonists in medium 199 containing 10% supplemented calf serum or fetal calf serum (FCS) (HyClone). OxPAPC and other lipids were added to cells 20 min before the addition of stimuli. Longer preincubation did not produce stronger inhibition of the effects of LPS. In the absence of serum, the cells were insensitive to LPS.

Cell ELISA

For the enzyme-linked immunosorbent assay (ELISA), HUVEC were treated with lipids and agonists for 4 h, and then the cells were washed and fixed. Cell-surface-expressed E-selectin, ICAM-1 or VCAM-1 was detected using corresponding antibodies (R&D

Systems), secondary peroxidase-conjugated antibodies and *o*-phenylenediamine as substrate.

Quantification of NF- κ B-dependent transcription

Co-transfection of HUVEC with pNF- κ B-Luc (Stratagene) and pCMV β (Clontech) was performed as described¹⁰. Two days after transfection, the cells were stimulated with 300 ng ml⁻¹ LPS in the presence or absence of OxPAPC, and 6 h later were processed for luciferase and β -galactosidase activity measurements.

Analysis of p65/DNA binding

HUVEC were stimulated with 300 ng ml⁻¹ LPS for 60 min in the presence or absence of 25 μ g ml⁻¹ OxPAPC. Analysis of p65 binding to its consensus oligonucleotide was performed in HUVEC lysates using the ELISA-based Trans-AM NF- κ B p65 kit (Active Motif).

Western blotting

HUVEC were stimulated with 300 ng ml⁻¹ LPS or 20 U ml⁻¹ TNF α in the presence or absence of 50 μ g ml⁻¹ OxPAPC. After 20 min (phospho-I κ B α), 1 h (total I κ B α) or 2 h (p38 MAPK), cells were scraped and analysed by western blotting as described¹⁰.

Limulus amoebocyte lysate assay

OxPAPC aliquots (60 μ g ml⁻¹) were mixed with equal volumes of LPS dilutions to produce final concentrations indicated on the x-axis in Fig. 2b. After 5 min at 37 °C, LAL reagent was added and LAL activity was determined by a quantitative chromogenic assay (Endochrome, Charles River Endosafe).

Binding of biotin-LPS to CD14 and LBP

LPS was biotinylated using EZ-Link Biotin-LC-Hydrazide (Pierce). Binding of biotin-LPS to recombinant human CD14 (Biometec) adsorbed onto ELISA plates was measured as described in ref. 24. Binding reaction was performed for 30 min at 37 °C in PBS/0.5% BSA buffer containing 1% FCS. In the absence of serum, binding of biotin-LPS to CD14 was below the detection limit. Bound biotin-LPS was detected with streptavidin-peroxidase and *o*-phenylenediamine. Binding of biotin-LPS to recombinant LBP (Biometec) captured by anti-LBP (HM14, a gift from W. A. Buurman) adsorbed to ELISA plates was measured as described in ref. 25. The binding was performed in PBS/0.1% BSA buffer. Detection of bound biotin-LPS was performed as described for CD14.

Animals

Experiments were performed according to Austrian animal rights law using female C57BL/6J mice (Institut für Labortierkunde und -genetik) of 20 g body weight.

Skin model of inflammation

LPS (10 μ g) with or without OxPAPC (50 μ g) in the final volume of 50 μ l of 0.9% saline was injected intradermally into the abdominal region. After 24 h, mice were killed, and the skin at the injection site was excised and either embedded in optimal-cutting-temperature compound for immunohistochemistry or homogenized in potassium phosphate buffer (50 mM, pH 6.0) containing 0.5% hexadecyltrimethylammonium bromide. Sections of skin were stained for myeloperoxidase (MPO) using anti-MPO (from Dako) and Alexa 466-labelled secondary antibodies (Molecular Probes). Nuclei were stained with DAPI (Sigma-Aldrich). MPO activity in homogenates was monitored at 460 nm using H₂O₂ and *o*-dianisidine (Sigma-Aldrich) as substrates²⁶.

Air pouch model of inflammation

Air pouches of 5 ml were raised on day 0 on the dorsal surface of mice under light halothane anaesthesia, and a further 3 ml of air was injected 3 days later²³. At day seven, 50 μ g of LPS, 250 μ g of OxPAPC, or a combination of both, was injected into the pouch in 1 ml of 0.9% saline. Twenty-four hours later, mice were killed, the pouches were lavaged and leukocytes were counted. Sections of air pouch wall were stained for p65 (Santa Cruz Biotechnology) or VCAM-1 (Southern Biotechnology Associates). Nuclei were stained with DAPI.

Peritoneal model of inflammation

Mice were injected intraperitoneally (i.p.) with 10 μ g of LPS with or without 50 μ g of OxPAPC at days 0 and 1. At day 2, the animals were killed and peritoneal leukocytes were counted in the lavage fluid. Paraffin-embedded sagittal sections of the anterior abdominal wall were stained with haematoxylin and eosin.

Survival studies

Mice were injected i.p. with 0.6 mg of LPS with or without 1.2 mg of OxPAPC in 0.5 ml of 0.9% saline. Survival was checked three times daily for up to 5 days.

mRNA expression

The mice were killed, and isolated organs were immersed into the RNAlater solution (Ambion). One μ g of total RNA isolated using the RNeasy lysis kit (Qiagen) was reverse transcribed using GeneAmp RNA PCR core kit (Applied Biosystems).

Quantification of E-selectin mRNA was performed by real-time polymerase chain reaction (PCR) using the LightCycler instrument and Fast Start DNA Master SYBR Green I kit (Roche Diagnostics) as recommended by the manufacturer. The following primers were used: forward, 5'-GGTTTGGTGAGGTGTGCTC; reverse, 5'-TGATCTTTCCCGGAAC TGC. Each sample was also analysed for expression of β_2 -microglobulin, which was used for normalization of E-selectin data.

Statistics

The data are presented as mean values $\pm$ standard deviations. The data are representative of at least two experiments producing similar results.

Received 30 May; accepted 10 July 2002; doi:10.1038/nature01023.

- Medzhitov, R. Toll-like receptors and innate immunity. *Nature Rev. Immunol.* **1**, 135–145 (2001).
- Aderem, A. & Ulevitch, R. J. Toll-like receptors in the induction of the innate immune response. *Nature* **406**, 782–787 (2000).
- Chow, J. C., Young, D. W., Golenbock, D. T., Christ, W. J. & Gusovsky, F. Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J. Biol. Chem.* **274**, 10689–10692 (1999).
- Beutler, B. Tlr4: central component of the sole mammalian LPS sensor. *Curr. Opin. Immunol.* **12**, 20–26 (2000).
- Hampton, M. B., Kettle, A. J. & Winterbourn, C. C. Inside the neutrophil phagosome: oxidants, myeloperoxidase, and bacterial killing. *Blood* **92**, 3007–3017 (1998).
- Zhang, R., Shen, Z., Nauseef, W. M. & Hazen, S. L. Defects in leukocyte-mediated initiation of lipid peroxidation in plasma as studied in myeloperoxidase-deficient subjects: systematic identification of multiple endogenous diffusible substrates for myeloperoxidase in plasma. *Blood* **99**, 1802–1810 (2002).
- Lusis, A. J. Atherosclerosis. *Nature* **407**, 233–241 (2000).
- Leitinger, N. *et al.* Structurally similar oxidized phospholipids differentially regulate endothelial binding of monocytes and neutrophils. *Proc. Natl Acad. Sci. USA* **96**, 12010–12015 (1999).
- Watson, A. D. *et al.* Structural identification by mass spectrometry of oxidized phospholipids in minimally oxidized low density lipoprotein that induce monocyte/endothelial interactions and evidence for their presence *in vivo*. *J. Biol. Chem.* **272**, 13597–13607 (1997).
- Bochkov, V. N. *et al.* Oxidized phospholipids stimulate tissue factor expression in human endothelial cells via activation of ERK/EGR-1 and Ca⁺⁺/NFAT. *Blood* **99**, 199–206 (2002).
- Elass-Rochard, E. *et al.* Lactoferrin inhibits the endotoxin interaction with CD14 by competition with the lipopolysaccharide-binding protein. *Infect. Immun.* **66**, 486–491 (1998).
- Emancipator, K., Csako, G. & Elin, R. J. *In vitro* inactivation of bacterial endotoxin by human lipoproteins and apolipoproteins. *Infect. Immun.* **60**, 596–601 (1992).
- Kitchens, R. L., Thompson, P. A., Viriyakosol, S., O'Keefe, G. E. & Munford, R. S. Plasma CD14 decreases monocyte responses to LPS by transferring cell-bound LPS to plasma lipoproteins. *J. Clin. Invest.* **108**, 485–493 (2001).
- Wurfel, M. M. & Wright, S. D. Lipopolysaccharide-binding protein and soluble CD14 transfer lipopolysaccharide to phospholipid bilayers: preferential interaction with particular classes of lipid. *J. Immunol.* **158**, 3925–3934 (1997).
- Rustici, A. *et al.* Molecular mapping and detoxification of the lipid A binding site by synthetic peptides. *Science* **259**, 361–365 (1993).
- Christ, W. J. *et al.* E5531, a pure endotoxin antagonist of high potency. *Science* **268**, 80–83 (1995).
- Golenbock, D. T., Hampton, R. Y., Qureshi, N., Takayama, K. & Raetz, C. R. Lipid A-like molecules that antagonize the effects of endotoxins on human monocytes. *J. Biol. Chem.* **266**, 19490–19498 (1991).
- Tanamoto, K. & Azumi, S. Salmonella-type heptaacylated lipid A is inactive and acts as an antagonist of lipopolysaccharide action on human line cells. *J. Immunol.* **164**, 3149–3156 (2000).
- Frey, E. A. *et al.* Soluble CD14 participates in the response of cells to lipopolysaccharide. *J. Exp. Med.* **176**, 1665–1671 (1992).
- Su, G. L., Simmons, R. L. & Wang, S. C. Lipopolysaccharide binding protein participation in cellular activation by LPS. *Crit. Rev. Immunol.* **15**, 201–214 (1995).
- Sugiyama, T. & Wright, S. D. Soluble CD14 mediates efflux of phospholipids from cells. *J. Immunol.* **166**, 826–831 (2001).
- Yu, B., Hailman, E. & Wright, S. D. Lipopolysaccharide binding protein and soluble CD14 catalyze exchange of phospholipids. *J. Clin. Invest.* **99**, 315–324 (1997).
- Lawrence, T., Gilroy, D. W., Colville-Nash, P. R. & Willoughby, D. A. Possible new role for NF- κ B in the resolution of inflammation. *Nature Med.* **7**, 1291–1297 (2001).
- Cunningham, M. D. *et al.* *Helicobacter pylori* and *Porphyromonas gingivalis* lipopolysaccharides are poorly transferred to recombinant soluble CD14. *Infect. Immun.* **64**, 3601–3608 (1996).
- Scott, M. G., Vreugdenhil, A. C., Buurman, W. A., Hancock, R. E. & Gold, M. R. Cutting edge: cationic antimicrobial peptides block the binding of lipopolysaccharide (LPS) to LPS binding protein. *J. Immunol.* **164**, 549–553 (2000).
- Bradley, P. P., Priebe, D. A., Christensen, R. D. & Rothstein, G. Measurement of cutaneous inflammation: estimation of neutrophil content with an enzyme marker. *J. Invest. Dermatol.* **78**, 206–209 (1982).

Acknowledgements

This project was funded by the Austrian Science Foundation and by the ICP Program of the Austrian Federal Ministry for Education, Science and Culture.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.L. (e-mail: norbert.leitinger@univie.ac.at).

Actin dynamics in the contractile ring during cytokinesis in fission yeast

Robert J. Pelham Jr & Fred Chang

Department of Microbiology, Columbia University College of Physicians and Surgeons, 701 West 168th Street, New York, New York 10032, USA

Cytokinesis in many eukaryotes requires a contractile ring of actin and myosin that cleaves the cell in two. Little is known about how actin filaments and other components assemble into this ring structure and generate force^{1,2}. Here we show that the contractile ring in the fission yeast *Schizosaccharomyces pombe* is an active site of actin assembly. This actin polymerization activity requires Arp3, the formin Cdc12, profilin and WASP, but not myosin II or IQGAP proteins. Both newly polymerized actin filaments and pre-existing actin cables can contribute to the initial assembly of the ring. Once formed, the ring remains a dynamic structure in which actin and other ring components continuously assemble and disassemble from the ring every minute. The rate of actin polymerization can influence the rate of cleavage. Thus, actin polymerization driven by the Arp2/3 complex and formins is a central process in cytokinesis. Our studies show that cytokinesis is a more dynamic process than previously thought and provide a perspective on the mechanism of cell division.

Genetic screens have identified genes that are required for contractile ring assembly in fission yeast³. One of these genes, *cdc12*, encodes a formin from a conserved family of proteins (including Bni1, For3, Diaphanous and mDia) that is required for the assembly of actin structures, such as contractile rings and actin cables for cytokinesis and cell polarity^{4,5}. Cdc12 localizes to the contractile ring, functions specifically in ring assembly and interacts with profilin Cdc3 (ref. 6), an actin-binding protein that is also essential for ring assembly⁷. Although the Arp2/3 complex, a well known nucleator of actin filaments, is required for the assembly of numerous actin structures^{8,9}, the functions of Arp2/3 and actin polymerization in cytokinesis have not been determined.

We examined whether the Arp2/3 complex is present in the *S. pombe* contractile ring using a functional fusion of green fluorescent protein and one of the Arp2/3 complex components, Arc15 (ref. 9). Arc15-GFP was localized in actin patches and in a medial ring structure that colocalized with the F-actin ring (Fig. 1 and Supplementary Information Movie 1). Immunofluorescence studies in fixed cells have shown that Arp2 and Arp3 are distributed in a diffuse medial pattern, but it is possible that the more discrete ring pattern may be lost on fixation^{10,11}. Although GFP fusions with some components of the Arp2/3 complex are not functional (K. Gould, personal communication), other proteins that are thought to be associated with the Arp2/3 complex, including Crn1 (coronin), App1 and Myo1, are also present in the contractile ring (refs 12–14; and B. Goode, personal communication).

We next determined whether the contractile ring functions as a site of actin polymerization. We used an *in vitro* actin polymerization assay in which permeabilized fission yeast cells were incubated with rhodamine-labelled G-actin and then assayed by fluorescence microscopy for incorporation of rhodamine fluorescence into the ring¹². These cells expressed Crn1-GFP as a marker for the contractile ring and actin patches. In wild-type cells, rhodamine-actin incorporated into actin patches¹² and into a discrete medial ring structure that colocalized with the Crn1-GFP ring (Fig. 2a, b). Incorporation occurred even in the initial stages of ring assembly, but as a more diffuse medial ring (Fig. 2a, top, upper right cell). Addition of 50 μ M Latrunculin A (Lat A), an inhibitor of actin

polymerization¹⁵, inhibited the incorporation of rhodamine-actin, which showed that incorporation requires actin polymerization (Fig. 2a, b).

Incorporation occurred at normal rates in permeabilized cells pretreated with a pulse of 200 nM cytochalasin D, which caps actin barbed ends¹⁶, indicating that incorporation did not arise from polymerization occurring at the barbed ends or at the slow-growing pointed ends of pre-existing filaments (Fig. 2a–c and data not shown). When preformed rhodamine-actin filaments were added to permeabilized cells, they were not incorporated in any specific cellular structure (data not shown), which suggests that *in vitro* the ring may not capture actin filaments assembled at other cellular structures. Together, these properties indicate that the ring contains an activity that promotes actin nucleation and polymerization.

Rhodamine-actin incorporation was greatly reduced in cold-sensitive *arp3-c1* cells (Fig. 2a–c). Rhodamine-actin incorporation was also inhibited by adding an antibody against Arp3 to permeabilized *arp3*⁺ cells, but not a control antibody, which suggested that Arp3 might function in actin polymerization at the ring (data not shown). We also examined the role of Wsp1 (WASP), an activator of Arp2/3 complex that shares a partial redundant function with another Arp2/3 complex activator, Myo1 (Myosin I)¹³. *wsp1* Δ mutants showed a partial loss of actin polymerization activity (Fig. 2a–c). Thus, some actin polymerization activity at the ring is dependent on Arp3.

We tested the activity of other proteins implicated in actin ring assembly by assaying temperature-sensitive actin ring mutants in

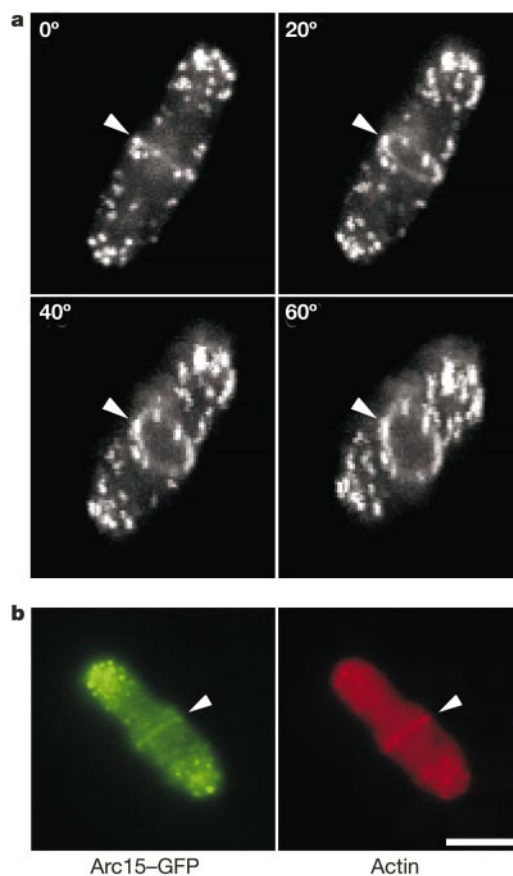


Figure 1 Arc15, a component of the Arp2/3 complex, localizes to the contractile ring. **a**, Three-dimensional confocal images of an Arc15-GFP medial ring (arrowhead), rotated at 20° intervals. **b**, Cell expressing Arc15-GFP (green) stained for F-actin (rhodamine phalloidin, red). Scale bar, 5 μ m.

two ways: in *in vivo* temperature-shift assays, cells were grown at restrictive temperature before permeabilization; in *in vitro* temperature-shift assays, cells were grown at permissive temperature, permeabilized and then assayed for activity at restrictive temperature. After the *in vivo* temperature shift, *cdc12-112* (formin) and *cdc3-313* (profilin) mutants showed no ring structures and no mitosis-specific medial actin polymerization activity (Fig. 2). After the *in vitro* temperature shift, *cdc3-313* and *cdc12-112* mutant cells contained GFP-labelled rings but had significantly reduced actin polymerization activity in the rings (Fig. 2). In *cdc3-313* mutants, actin polymerization activity was absent in both actin rings and patches (Fig. 2). In the *cdc12-112* mutant, actin polymerization activity was not eliminated in surrounding actin patches, which is consistent with the specific localization and function of Cdc12 in the ring⁶. The *in vitro* temperature sensitivity of the *cdc12* and *cdc3* mutants indicates that Cdc12 and Cdc3, which are essential for ring assembly, may function directly in actin polymerization and/or nucleation at the contractile ring.

By contrast, other proteins required for ring assembly, Cdc4 (a myosin light chain), Myo2 (myosin II) and Rng2 (IQGAP), were not needed for this activity. After an *in vivo* temperature shift, *cdc4*, *myo2* and *rng2* mutants, which form abnormal actin structures instead of a ring^{17,18}, showed one or more medial foci of actin

polymerization that were distinct from actin patches (data not shown). On *in vitro* temperature shift, actin polymerization in the rings of these mutants was not affected (Supplementary Information).

Next, we determined whether the contractile ring is an active site of F-actin polymerization *in vivo*. The contractile ring persists in mitosis (for 20–30 min) with a steady-state amount of F-actin present in the ring. Thus, the rate of actin polymerization can be determined by the rate of F-actin turnover. When actin polymerization was inhibited by a high dose (200 μ M) of Lat A, F-actin in the ring disappeared in 60 s (Fig. 3a, b) and no stable forms of the ring were detected. These data therefore suggest that F-actin turnover in the contractile ring occurs in roughly 1 min, similar to rates seen in actin patches and cables¹².

We next determined whether formin Cdc12 and profilin Cdc3 are required for this continuous actin polymerization *in vivo*. When temperature-sensitive *cdc3-313* and *cdc12-112* mutants were shifted to restrictive temperature for 10 min, actin rings were not maintained (Fig. 3c, d). Thus, maintenance of the ring requires F-actin polymerization that is dependent on Cdc12 and Cdc3.

We used fluorescence recovery after photobleaching (FRAP) to determine the dynamics of ring components. Although we could not examine actin exchange directly in this assay because of the lack of a functional GFP–actin fusion protein, we examined dynamics of

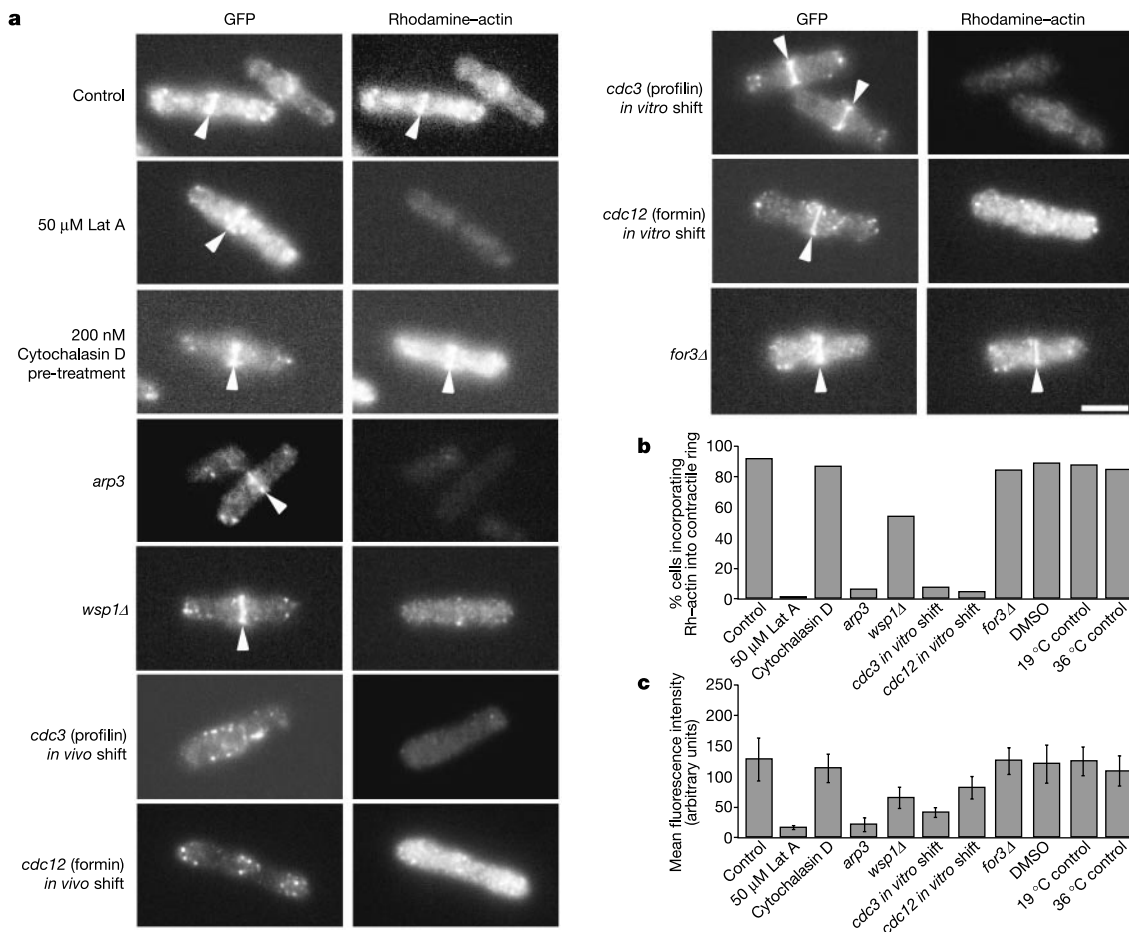


Figure 2 The contractile ring is a site of actin polymerization and/or nucleation *in vitro*. Cells of the indicated genotype expressing Crn1–GFP were permeabilized and incubated with rhodamine-labelled G-actin. Actin polymerization activity was measured by the incorporation of rhodamine fluorescence. **a**, Images of permeabilized cells in the assay. Rhodamine fluorescence (right) indicates rhodamine-actin incorporation into the

contractile ring (arrowheads) and GFP fluorescence (left) from Crn1–GFP. **b**, Percentage of cells incorporating rhodamine-actin into the contractile ring ($n > 81$). **c**, Mean fluorescence intensity of rhodamine-actin incorporation in the contractile ring in each cell. Scale bar, 5 μ m.

two other ring proteins: Cdc4 (a myosin light chain); and Cdc8 (tropomyosin), a protein that binds to the sides of actin filaments¹⁹. Both GFP-Cdc4 and GFP-Cdc8 fluorescence recovered rapidly after photobleaching, with $\tau_{1/2} < 30$ s (mean $\pm$ s.d., 27 ± 1.3 s, $n = 5$ cells, and 24.4 ± 1.0 s, $n = 5$ cells, respectively; Fig. 3e–g). Similar recovery times were found in closing rings. Rings that were completely photobleached recovered with similar kinetics (data not shown). Thus, several components of the actin ring exchange on and off the ring rapidly *in vivo*.

We examined the mechanism of actin ring assembly. Studies in cells from mammals, *Dictyostelium* and fission yeast suggest that pre-existing actin cables from other regions of the cell may be transported to the medial region and reorganized to form the contractile ring^{20–22}. Our studies suggest an alternative model in which the actin ring may also be formed by *de novo* actin poly-

merization. To test whether interphase actin cables are necessary for actin ring assembly, we examined a *for3Δ* (another formin) mutant that lacks interphase actin cables²³. In early stages of ring formation, wild-type cells form a medial web or loose ring of actin cables that seems to be contiguous with longitudinal actin cables (Fig. 3h, WT arrowhead)²². *for3Δ* cells had no detectable interphase actin cables or medial web of actin cables; however, they did show well-defined medial actin rings, even in early mitosis (Fig. 3h, *for3Δ*, arrowhead). Actin polymerization activity was normal in *for3Δ* rings (Fig. 2a–c). Thus, the ring can assemble in the absence of detectable interphase actin cables.

Actin polymerization is required for ring assembly because rings are absent in cells treated with Lat A and in *cdc12* or *cdc3* mutants (Fig. 3). But the function of the Arp2/3 complex in ring assembly is less clear. In budding yeast, actin-ring-like structures form in strains

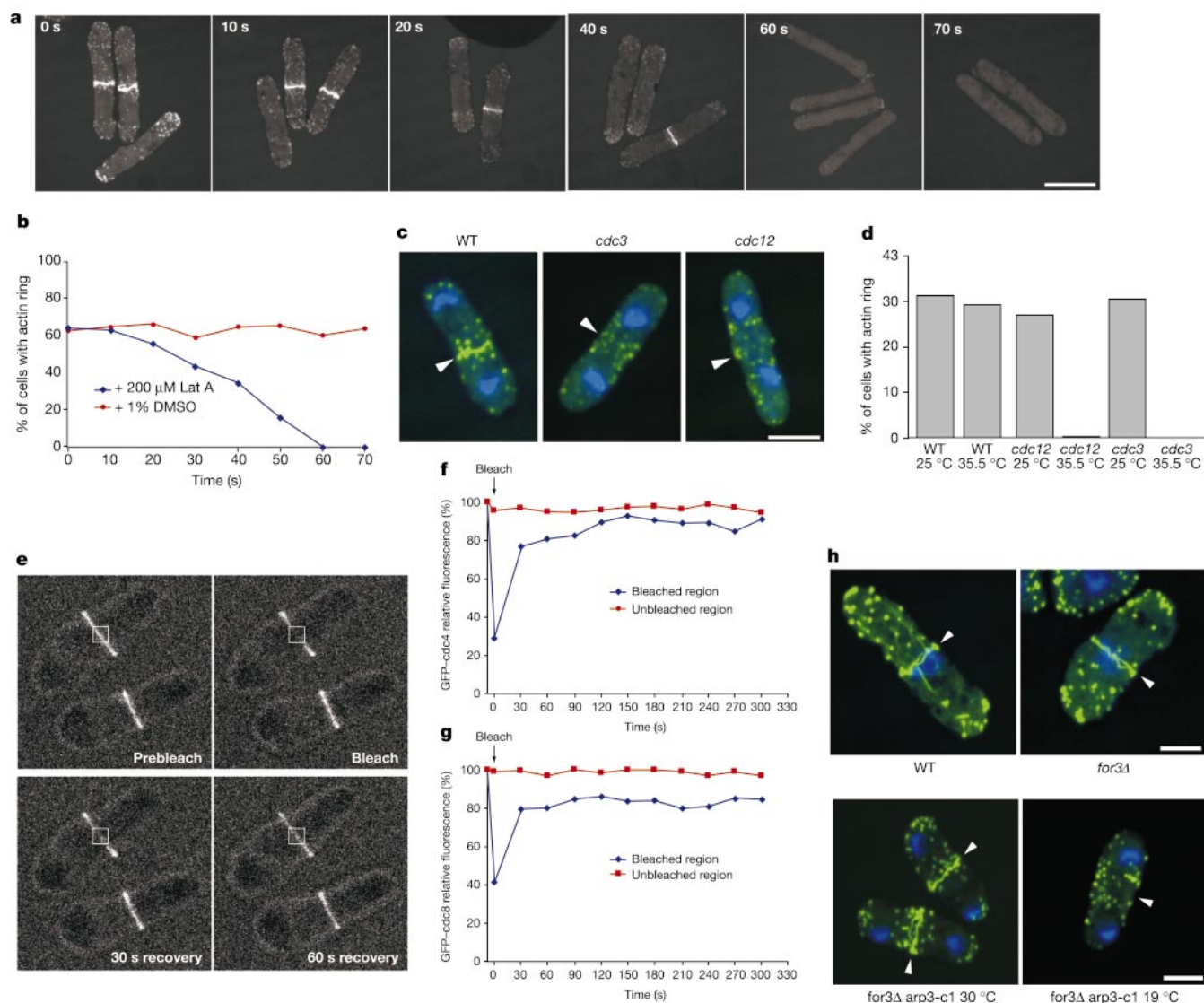


Figure 3 Actin and other components of the contractile ring are dynamic *in vivo*. **a, b**, Turnover of F-actin in the contractile ring *in vivo*. **a**, Mitotic cells treated with 200 μ M Lat A stained for F-actin with AlexaFluor 488 phalloidin. **b**, Kinetics of actin ring loss after treatment with Lat A ($n = 624$ cells). **c, d**, Maintenance of F-actin in the ring requires profilin and formin activity. **c**, Wild-type (WT), temperature-sensitive *cdc3-313* (profilin) and *cdc12-112* (formin) cells stained with AlexaFluor 488 phalloidin (green) and 4',6-diamidino-2-phenylindole dihydrochloride (DAPI; blue) after a 10-min temperature shift. **d**, Percentage of cells with an actin ring after disruption of Cdc3 (profilin) or Cdc12 (formin)

activity ($n > 30$). **e–g**, Dynamics of the ring components Cdc4 and Cdc8. **e**, Representative images of cells expressing GFP-Cdc4 in FRAP experiments. Box indicates area of contractile ring that was photobleached. **f**, Recovery of GFP-Cdc4 fluorescence in the ring after photobleaching. **g**, Recovery of GFP-Cdc8 fluorescence in the ring after photobleaching. **h**, Wild-type, *for3Δ* and *for3Δ arp3-c1* cells (at both permissive and restrictive temperatures) were fixed and stained for actin (AlexaFluor 488 phalloidin, green) and DNA (DAPI, blue). Scale bars, 5 μ m.

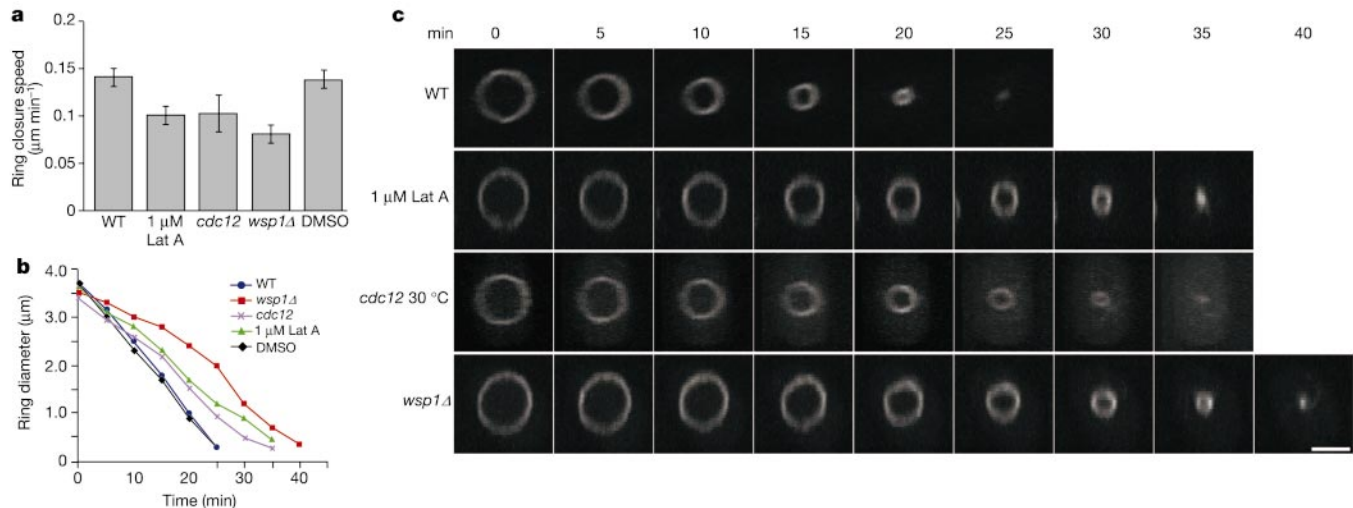


Figure 4 Actin polymerization influences the rate of contractile ring closure. Analysis of contractile ring closure in wild-type, *wsp1Δ* and *cdc12-112* cells and in wild-type cells treated with 1 μM Lat A or DMSO at 30 °C. **a**, Rates of contractile ring closure ($n > 12$). **b**,

Kinetics of contractile ring closure. **c**, Time-lapse images of GFP-Cdc4 rings. Scale bar, 2.5 μm.

that lack subunits of the Arp2/3 complex²⁴. In fission yeast, actin rings are still present in cold-sensitive *arp3-c1* mutant cells²⁵, although this conditional allele may confer only a partial loss of Arp2/3 function. To test whether both actin polymerization and actin cables contribute to ring assembly, we constructed a *for3Δ arp3-c1* double mutant. *for3Δ arp3-c1* cells failed to form actin rings at restrictive temperature (Fig. 3h, *for3Δ arp3-c1*, 19 °C, arrowhead), whereas this temperature shift had little effect on actin rings in wild-type, *arp3-c1* or *for3Δ* cells (data not shown). It is possible that in the absence of actin cables, normal levels of Arp3 activity are required for ring assembly. Thus, formin-dependent actin cables and Arp3-dependent actin polymerization share some functional redundancy and both may contribute to actin ring assembly.

We next determined whether actin polymerization contributes to the function of the contractile ring in cell cleavage. As complete inhibition of actin polymerization causes rapid loss of the contractile ring (Fig. 3), we examined ring closure in three cell populations in which actin polymerization was attenuated: (1) cells treated with 1 μM Lat A; (2) *wsp1Δ* mutant cells; and (3) *cdc12-112* mutant cells at semi-permissive temperature (30 °C). In these cells the rings had normal morphology, as determined by actin staining (data not shown) and GFP-Cdc4 fluorescence (Fig. 4c), and closed in a coordinated manner (Fig. 4b), but the rates of ring closure were significantly slower than those of the wild type (Fig. 4). Thus, a reduction in actin polymerization leads to a decrease in the rate of ring closure.

As the movement of actin patches also depends on actin polymerization¹², we compared the rates of actin patch movement with the rates of ring closure. As compared with wild-type cells, the rates of actin patch movement in *wsp1Δ* cells and in cells treated with 1 μM Lat A were decreased by 44% and 31%, and the rates of ring closure were decreased by 43% and 34%, respectively (Supplementary Information Table 1). These similar reductions suggest that the movements of the actin patch and contractile ring may involve related actin polymerization-dependent mechanisms.

It is not clear how actin dynamics might contribute to ring closure. As Cdc12 is found only in the contractile ring, the effect of a *cdc12* mutation in ring closure might arise from the attenuation of actin polymerization specifically at the ring and not elsewhere in the cell. Actin dynamics might contribute to septum formation and

membrane insertion at the ring, or might influence interactions between actin and myosin. As actin polymerization can generate force²⁶, another possibility is that actin polymerization in the contractile ring might contribute force for cleavage.

In summary, we have shown that the contractile ring is a dynamic structure, with actin and its other components exchanging every minute. This is in contrast to a previous view that the ring consists of relatively stable actin filaments that gradually slide together from myosin-based forces¹. The active polymerization (and depolymerization) of actin filaments can contribute to the assembly, maintenance and closure of the contractile ring. Both the formin Cdc12 and the Arp2/3 complex are required for actin polymerization activity; however, our studies do not distinguish whether formins function upstream, downstream, or in a separate pathway from Arp2/3 (refs 27, 28). Formins and profilin have also been implicated in cytokinesis in animal cells⁴, and labelled G-actin rapidly incorporates into the *Xenopus* cleavage furrow²⁹. Thus, it is likely that actin polymerization is important for cytokinesis in many other eukaryotes. □

Methods

Yeast strains, media and genetic methods

The *S. pombe* strains used are listed in Supplementary Information Table 2. Standard methods for media and genetic manipulations are described (<http://www.bio.uva.nl/pombe>). The Arc15-GFP strain (a gift from K. Gould) contains a GFP-kanMX cassette targeted to the chromosomal *arc15⁺* locus and expresses Arc15-GFP from the endogenous *arc15⁺* promoter. pREP42-GFP-*cdc4* (ref. 30) and pREP42-GFP-*cdc8* (M. Balasubramanian, personal communication) were gifts from M. Balasubramanian. We carried out plasmid transformations using the Frozen-EZ Yeast Transformation II kit (Zymo Research).

Microscopy and pharmacological inhibitors

We carried out microscopy, phalloidin staining and analysis of actin patch movement as described¹². Stock solutions of Lat A (provided by D. Drubin) and cytochalasin D (Sigma) were prepared in dimethyl sulphoxide (DMSO). For the experiments shown in Fig. 3a and b, *cdc25-22* cells were grown at 36 °C for 4 h, shifted to 25 °C and monitored for about 100 min until enriched for mitotic cells.

We carried out FRAP experiments on a Zeiss LSM 510 confocal microscope. Fluorescence recovery was analysed using Zeiss 3D for LSM software, and $\tau_{1/2}$ was calculated as the time needed to reach half of the final intensity after photobleaching. Unbleached rings in the same field were used as controls. This photobleaching protocol did not perturb the function of the ring, because cells completed cytokinesis after photobleaching. Four-dimensional images of GFP-Cdc4 in the contractile ring were obtained using a spinning-disk confocal microscope (Perkin Elmer-Wallac)¹². We deconvolved and processed z-sections of 35 × 0.2 μm acquired at 5-min intervals using Openlab software (Improvision).

In vitro actin polymerization assay

The *in vitro* actin polymerization assay was done as described¹² using rhodamine-labelled rabbit skeletal muscle actin (Cytoskeleton). For *in vivo* shift experiments with the temperature-sensitive strains, cells were shifted to 36°C for 180 min before and during collection. In *in vitro* shift assays, cells were grown at 25°C (permissive temperature), permeabilized, shifted to 36°C (restrictive temperature) for 30 min *in vitro*, and then kept at this temperature during incubation with rhodamine-actin. Cold-sensitive *arp3-c1* cells were shifted to 19°C for 30 min before collection and kept at this temperature during collection. Cells were treated with 50 µM Lat A or 1% DMSO during rhodamine-actin incorporation. Cells were treated with 200 nM cytochalasin D after permeabilization and then with 20 nM cytochalasin D during rhodamine-actin incorporation¹⁶. To test whether F-actin incorporates into permeabilized cells, we added pre-assembled rhodamine-actin filaments¹⁴ to permeabilized cells with and without 10 µM phalloidin (Molecular Probes).

Received 1 May; accepted 27 June 2002; doi:10.1038/nature00999.

- Rappaport, R. *Cytokinesis in Animal Cells* 168–184 (Cambridge Univ. Press, Cambridge, 1996).
- Robinson, D. N. & Spudich, J. A. Towards a molecular understanding of cytokinesis. *Trends Cell Biol.* **10**, 228–237 (2000).
- Feierbach, B. & Chang, F. Cytokinesis and the contractile ring in fission yeast. *Curr. Opin. Microbiol.* **4**, 713–719 (2001).
- Wasserman, S. FH proteins as cytoskeletal organizers. *Trends Cell Biol.* **8**, 111–115 (1998).
- Sawin, K. E. Cell polarity: following formin function. *Curr. Biol.* **12**, R6–R8 (2002).
- Chang, F., Drubin, D. & Nurse, P. cdc12p, a protein required for cytokinesis in fission yeast, is a component of the cell division ring and interacts with profilin. *J. Cell Biol.* **137**, 169–182 (1997).
- Balasubramanian, M. K., Hirani, B. R., Burke, J. D. & Gould, K. L. The *Schizosaccharomyces pombe* *cdc3⁺* gene encodes a profilin essential for cytokinesis. *J. Cell Biol.* **125**, 1289–1301 (1994).
- Higgs, H. N. & Pollard, T. D. Regulation of actin filament network formation through Arp2/3 complex: activation by a diverse array of proteins. *Annu. Rev. Biochem.* **70**, 649–676 (2001).
- Machesky, L. M. & Gould, K. L. The Arp2/3 complex: a multifunctional actin organizer. *Curr. Opin. Cell Biol.* **11**, 117–121 (1999).
- Arai, R., Nakano, K. & Mabuchi, I. Subcellular localization and possible function of actin, tropomyosin and actin-related protein 3 (Arp3) in the fission yeast *Schizosaccharomyces pombe*. *Eur. J. Cell Biol.* **76**, 288–295 (1998).
- Morrell, J. L., Morpew, M. & Gould, K. L. A mutant of arp2p causes partial disassembly of the Arp2/3 complex and loss of cortical actin function in fission yeast. *Mol. Biol. Cell.* **10**, 4201–4215 (1999).
- Pelham, R. J. & Chang, F. Role of actin polymerization and actin cables in actin-patch movement in *Schizosaccharomyces pombe*. *Nature Cell Biol.* **3**, 235–244 (2001).
- Lee, W. L., Bezanilla, M. & Pollard, T. D. Fission yeast myosin-II, Myo1p, stimulates actin assembly by Arp2/3 complex and shares functions with WASp. *J. Cell Biol.* **151**, 789–800 (2000).
- Goode, B. L., Rodal, A. A., Barnes, G. & Drubin, D. G. Activation of the Arp2/3 complex by the actin filament binding protein Abp1p. *J. Cell Biol.* **153**, 627–634 (2001).
- Morton, W. M., Ayscough, K. R. & McLaughlin, P. J. Latrunculin alters the actin-monomer subunit interface to prevent polymerization. *Nature Cell Biol.* **2**, 376–378 (2000).
- Li, R., Zheng, Y. & Drubin, D. G. Regulation of cortical actin cytoskeleton assembly during polarized cell growth in budding yeast. *J. Cell Biol.* **128**, 599–615 (1995).
- Chang, F., Woollard, A. & Nurse, P. Isolation and characterization of fission yeast mutants defective in the assembly and placement of the contractile actin ring. *J. Cell Sci.* **109**, 131–142 (1996).
- Balasubramanian, M. K. *et al.* Isolation and characterization of new fission yeast cytokinesis mutants. *Genetics* **149**, 1265–1275 (1998).
- McCollum, D., Balasubramanian, M. K., Pelcher, L. E., Hemmingsen, S. M. & Gould, K. L. *Schizosaccharomyces pombe* *cdc4⁺* gene encodes a novel EF-hand protein essential for cytokinesis. *J. Cell Biol.* **130**, 651–660 (1995).
- Cao, L. G. & Wang, Y. L. Mechanism of the formation of contractile ring in dividing cultured animal cells. I. Recruitment of preexisting actin filaments into the cleavage furrow. *J. Cell Biol.* **110**, 1089–1095 (1990).
- Fukui, Y., Kitanishi-Yumura, T. & Yumura, S. Myosin II-independent F-actin flow contributes to cell locomotion in *Dictyostelium*. *J. Cell Sci.* **112**, 877–886 (1999).
- Arai, R. & Mabuchi, I. F-actin ring formation and the role of F-actin cables in the fission yeast *Schizosaccharomyces pombe*. *J. Cell Sci.* **115**, 887–898 (2002).
- Feierbach, B. & Chang, F. Roles of the fission yeast formin for3p in cell polarity, actin cable formation and symmetric cell division. *Curr. Biol.* **11**, 1656–1665 (2001).
- Winter, D. C., Choe, E. Y. & Li, R. Genetic dissection of the budding yeast Arp2/3 complex: a comparison of the *in vivo* and structural roles of individual subunits. *Proc. Natl Acad. Sci. USA* **96**, 7288–7293 (1999).
- McCollum, D., Feoktistova, A., Morpew, M., Balasubramanian, M. & Gould, K. L. The *Schizosaccharomyces pombe* actin-related protein, Arp3, is a component of the cortical actin cytoskeleton and interacts with profilin. *EMBO J.* **15**, 6438–6446 (1996).
- Theriot, J. A. The polymerization motor. *Traffic* **1**, 19–28 (2000).
- Sagot, I., Klee, S. K. & Pellman, D. Yeast formins regulate cell polarity by controlling the assembly of actin cables. *Nature Cell Biol.* **4**, 42–50 (2002).
- Evangelista, M., Pruyne, D., Amberg, D. C., Boone, C. & Bretscher, A. Formins direct Arp2/3-independent actin filament assembly to polarize cell growth in yeast. *Nature Cell Biol.* **4**, 32–41 (2002).
- Noguchi, T. & Mabuchi, I. Reorganization of actin cytoskeleton at the growing end of the cleavage furrow of *Xenopus* egg during cytokinesis. *J. Cell Sci.* **114**, 401–412 (2001).
- Balasubramanian, M. K., McCollum, D. & Gould, K. L. Cytokinesis in fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **283**, 494–506 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank K. Gould for the Arc15–GFP construct, Arp3 reagents and communicating results before publication; D. Burgess, B. Feierbach, D. McCollum, A. Paoletti, R. Kessin and B. Goode for comments on the manuscript; T. Swayne and the Optical Imaging Facility of the Herbert Irving Cancer Center for use of their confocal microscope; T. Pollard for the *wsp1Δ* strain; and M. Balasubramanian for the GFP–Cdc4 and GFP–Cdc8 plasmids. This work was supported by research grants to F.C. from the NIH and the American Cancer Society.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to F.C. (e-mail: fc99@columbia.edu).

The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing

Hansen Du & Michael Rosbash

Howard Hughes Medical Institute, Department of Biology, MS008, Brandeis University, 415 South Street, Waltham, Massachusetts 02454, USA

Splicing of precursor messenger RNA takes place in the spliceosome, a large RNA/protein macromolecular machine¹. Spliceosome assembly occurs in an ordered pathway *in vitro* and is conserved between yeast and mammalian systems. The earliest step is commitment complex formation in yeast or E complex formation in mammals, which engages the pre-mRNA in the splicing pathway and involves interactions between U1 small nuclear ribonucleoprotein (snRNP) and the pre-mRNA 5' splice site^{2,3}. Complex formation depends on highly conserved base pairing between the 5' splice site and the 5' end of U1 snRNA, both *in vivo* and *in vitro*^{4–7}. U1 snRNP proteins also contribute to U1 snRNP activity^{8–10}. Here we show that U1 snRNP lacking the 5' end of its snRNA retains 5'-splice-site sequence specificity. We also show that recombinant yeast U1C protein, a U1 snRNP protein, selects a 5'-splice-site-like sequence in which the first four nucleotides, GUAU, are identical to the first four nucleotides of the yeast 5'-splice-site consensus sequence. We propose that a U1C 5'-splice-site interaction precedes pre-mRNA/U1 snRNA base pairing and is the earliest step in the splicing pathway.

We previously showed that a truncated U1 snRNP, lacking the 5'-splice-site complementary 5' end of U1 snRNA, can form a pre-mRNA/U1 snRNP complex and still has some sequence specificity for a normal 5' splice site¹¹. To pursue this observation, we carried out *in vitro* selection (SELEX) experiments with U1 snRNP (Fig. 1, WT). We used a standard yeast commitment complex substrate in which the six-nucleotide 5' splice site (normally GUAUGU) had been randomized. The pool was incubated with a yeast extract under standard commitment complex conditions and then immunoprecipitated with an antibody against Prp40 (a U1 snRNP protein)⁸. After six rounds of selection, only the sequence GUAAGU was obtained. The selected sequence is predicted to form a more stable interaction with U1 snRNP than with a canonical GUAUGU 5' sequence, because the uracil to adenine change and the extra guanine of the selected sequence allows the formation of three additional base pairs.

To test whether the selection of the GUAAGU sequence was exclusively due to the 5' end of U1 snRNA, we repeated the selection

experiment in an extract treated with a complementary oligonucleotide and RNase H to digest that portion of the U1 molecule. Digestion was virtually complete and the U1 snRNA was 10 nucleotides shorter (data not shown)¹¹. With this truncated U1 snRNP, we were unable to identify a discrete, selected sequence under these same conditions (Fig. 1b, 25 °C). As the complexes that are formed with truncated U1 snRNP are less stable than wild-type complexes¹¹, this result was not unexpected. With incubation at 4 °C, however, the six-round selection protocol and RNA sequencing identified the same unique sequence, GUAAGU (Fig. 1); 9 out of 10 clones had this same sequence (data not shown). If this had been caused by small amounts of undigested U1 snRNP, we would not expect a temperature effect on the selection. The result suggests that one or more commitment complex proteins and/or another portion of U1 snRNA selects the same 5'-splice-site sequence. Similar results have been reported for mammalian U1 snRNP¹².

We focused on the U1 snRNP U1C protein because it is important in both mammalian and yeast splicing^{13–16} and cross-linking data indicate that U1C associates directly with the 5' splice site⁸. To test whether U1C also contacts the 5'-splice-site sequence in the absence of the canonical 5'-splice-site/U1 snRNA base pairing, we labelled unique 5'-splice-site phosphates adjacent to single 4-thiouracils in the pre-mRNA substrates. Standard cross-linking experiments were carried out on the truncated U1 snRNA extract (Fig. 2). The pattern was very similar to that observed with intact U1 snRNP⁸. U1C was the most intensely labelled protein, especially when crosslinked with RNAs that were labelled at phosphates between the 5'-splice-site A3 and U4 or between G5 and U6 (Fig. 2).

As the truncated U1 snRNP could successfully select a 5'-splice-

site sequence at low temperature, we compared complex formation efficiency at 25 °C and at 0 °C. In addition to the truncated U1 snRNP extract, we also tested a U1C-depleted extract and a wild-type extract (Fig. 3). With the modified extracts there was only a small, reproducible increase in efficiency at low temperature as compared with at 25 °C (Fig. 3a and data not shown). With a wild-type extract, however, there was a marked reduction in the amount of wild-type complex at low temperature (Fig. 3a). This effect could not be due to a decrease in complex stability, as this should be enhanced at low temperature. A more likely explanation, based on the lack of a comparable temperature effect in the truncated extract, is that complex formation is qualitatively different at low temperature and predominantly based on RNA/protein interactions rather than on the canonical RNA/RNA pairing interaction.

To test this possibility, we assayed 5'-splice-site/U1 snRNA base pairing by psoralen crosslinking (Fig. 3b). As previously shown, there were no bands in the truncated extract at either temperature¹¹. There were also no detectable bands in the wild-type extract at 0 °C, which is consistent with our hypothesis and indicates that there is little or no RNA/RNA base pairing (Fig. 3b, lanes 1 and 2). The same result was observed at 5 °C, and the normal crosslinked products were observed at 10 °C and above (Fig. 3c). In the U1C-depleted extract, the crosslinked bands were present at 0 °C and at 25 °C (Fig. 3b, lanes 4 and 5). The results suggest that wild-type U1 snRNP does not efficiently undergo 5'-splice-site/U1 snRNA duplex formation at low temperature, but base pairing occurs under these conditions in the absence of U1C.

To address directly RNA/protein interactions, we assayed the crosslinking of U1 snRNP proteins to a 4-thiouracil-labelled substrate (Fig. 3d). The standard crosslinking pattern was obtained with the wild-type extract at low temperature, although the U1C and Snup56p bands were more intense, and the SmD1 and SmD3 bands were less intense, than at 25 °C. Increases in U1C and Snup56 intensity were also apparent in the truncated extract, which suggests that there is enhanced interaction or stability of these RNA/protein interactions at low temperature even without the 5' end of U1 snRNA. In the U1C-depleted extract, the U1C band was noticeably less intense and the Snup56 band was absent, which is consistent

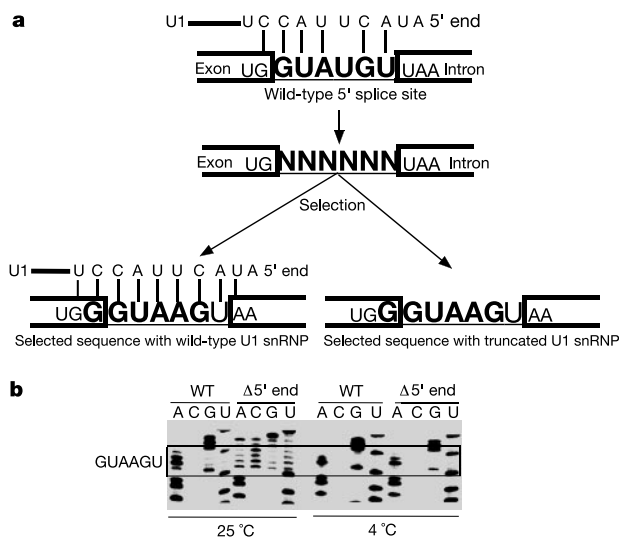


Figure 1 *In vitro* selection with U1 snRNP. **a**, Summary of selected sequences. The wild-type 5' splice site was randomized and used to select the preferred sequence for U1 snRNP binding, either with wild-type or truncated U1 snRNA. Bold letters indicate the selected sequences; large letters between brackets (GUAAGU) indicate the putative 5'-splice-site sequences. **b**, Winning sequences were determined by sequencing the double-stranded DNA pool after six rounds of selection. RNA pools were incubated in standard splicing extracts containing wild-type (WT) or truncated U1 snRNP, at either 25 °C or 4 °C. These two selections were done side by side simultaneously. Immunoprecipitated RNA was amplified by PCR with reverse transcription to generate an enriched double-stranded DNA pool, which was then sequenced. Ten individual clones from each pool were sequenced to confirm the winning sequence (data not shown).

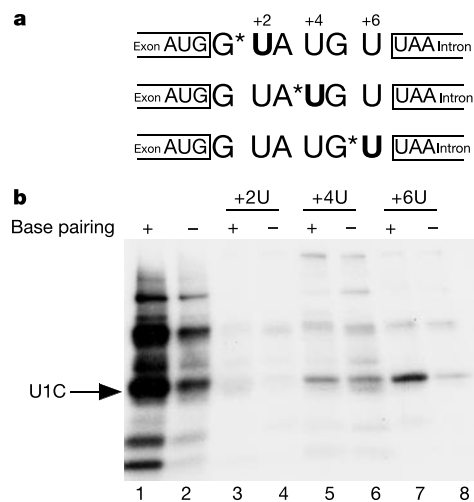


Figure 2 Site-specific crosslinking of commitment complex proteins to the 5'-splice-site region in the presence and absence of pre-mRNA/U1snRNA base pairing. **a**, Site-specifically labelled RNA substrates. Asterisks indicate ³²P-labelled positions. The 4-thiouracil is indicated in bold. **b**, Protein profiles that crosslink to the 5'-splice-site region. In lanes 1 and 2, body labelled RNA substrate was used with either wild-type or truncated U1 snRNP splicing extract¹¹, respectively. In lanes 3–8, site-specifically labelled substrates were used as indicated. The position of U1C is shown on the left.

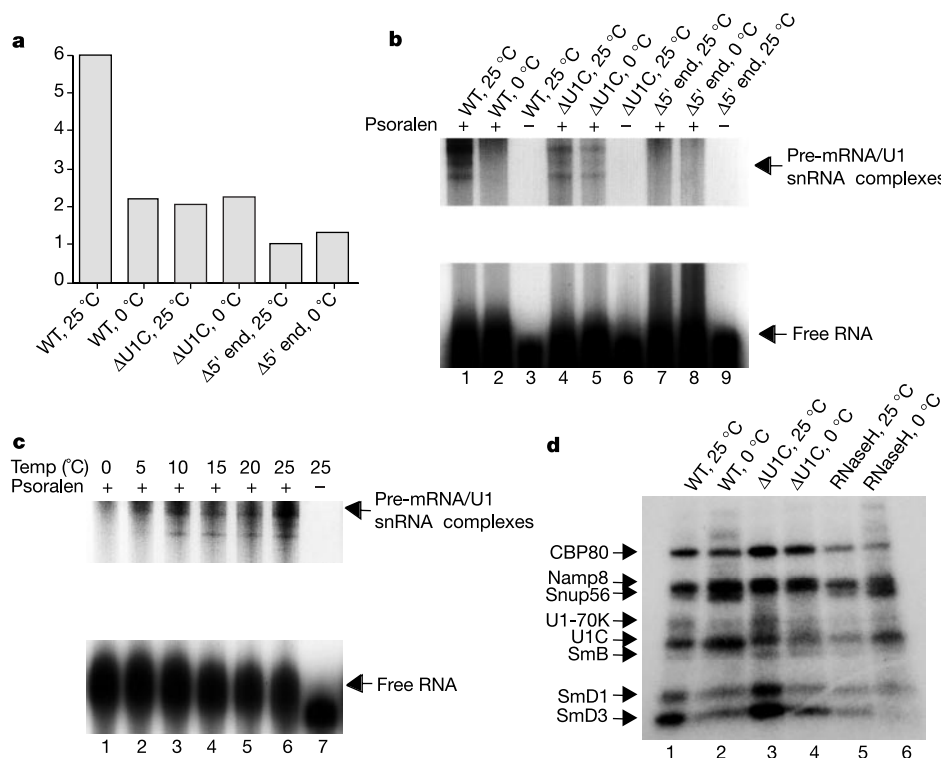


Figure 3 Characterization of complex formation in different extracts and at different temperatures. **a**, Quantitative analysis of complex formation. Three different splicing extracts were used to form RNA/protein complexes for 30 min with a standard body-labelled ^{32}P substrate of 72 nucleotides (WT-72)⁸. The reactions were at equilibrium as longer incubations had no effect, and identical results were obtained at 4 °C and at 0 °C (not shown). Complex formation was measured by immunoprecipitation with an antibody against Prp40 (ref. 11). **b**, U1 snRNA/5'-splice-site base pairing assayed by psoralen crosslinking. Crosslinking was assayed in the three extracts after 30 min of incubation at the two temperatures. Results and procedures for the wild-type and RNase-H-treated

extracts at 25 °C were as described¹¹. **c**, Psoralen crosslinking as a function of incubation temperature. The wild-type extract was incubated for 30 min at different temperatures, and base pairing was assayed by psoralen. **d**, Comparison of the protein crosslinking patterns of the three extracts at two different incubation temperatures. 4-Thiouracil-labelled ^{32}P WT-72 RNA⁸ was used as a substrate for protein ultraviolet crosslinking, and samples were normalized for counts before being loaded on the SDS gels so that similar amounts of complex were compared. Results for the wild-type and RNase-H-treated extract at 25 °C were as described^{8,11}. Previously identified proteins⁸ are indicated on the left.

with the idea that the U1C protein is missing from most of this U1 snRNP population and that the Snup56 interaction with RNA is dependent on U1C. The weak signal at the U1C position is due to residual U1C protein and/or to a co-migrating protein, which is presumably detectable when most of U1C is depleted.

Together, the results indicate that the predominant interaction at low temperature is between U1 snRNP proteins and the single-stranded 5'-splice-site region. This suggests that selection with wild-type U1 snRNP at 4 °C is based predominantly on RNA/protein interactions. Even at 25 °C, wild-type U1 snRNP selection might still be substantially due to RNA/protein interactions. This is because it is difficult to estimate the efficiency of canonical 5'-splice-site/U1 snRNA complex formation at 25 °C; in other words, there might still be a substantial fraction of RNA/protein complexes that have not converted to base pairing.

These results (Figs 2 and 3) suggest that U1C contributes to 5'-splice-site sequence specificity. We found that in electrophoretic mobility shift assays (EMSAs) recombinant yeast U1C could shift a substrate containing a normal 5' splice site but not one with several 5'-splice-site mutations (AUUUGU versus GUAUGU; Fig. 4a). To identify a preferred sequence, we carried out *in vitro* selection experiments using the RNA pool described above. In this case, however, filter binding was used to isolate RNAs that bound to recombinant U1C. After five rounds, the affinity was about 20-fold higher than with the initial pool (Fig. 4b). Only one winner sequence, GUAUAA, was apparent in the RNA sequence; it was also the only sequence in clones of this RNA (Fig. 4c). The first four

nucleotides are identical to a correctly positioned 5'-splice-site sequence, and the fifth nucleotide is a purine. (The yeast consensus sequence is a guanosine at this position.) This result does not exclude additional contacts between recombinant U1C and other regions of the 72-nucleotide RNA, which might contribute to the selection of the GUAU motif.

We presume that U1C interacts with the GUAUAA sequence as predominantly single-stranded RNA, because there are no more than three adjacent complementary bases in the substrate RNA (data not shown). To verify this notion, we carried out EMSA experiments with a partially double-stranded RNA. Before the addition of recombinant U1C protein, the winner substrate was preincubated with an antisense RNA roughly centred on the 5' splice site. There was no evidence for an interaction of U1C with the duplex RNA (Fig. 5). In addition, the results indicated that a double-stranded 5'-splice-site region competes with RNA/U1C complex formation.

Mutations of L13 in the conserved zinc-finger domain of U1C bypass the function of an essential helicase, Prp28 (ref. 10). It was proposed that the mutations decrease the stability of the U1 snRNA/5'-splice-site interaction, which then no longer requires the helicase for duplex melting and progression along the splicing pathway¹⁰. Previous results are also consistent with a duplex stabilization role for U1C^{13,14}. Although our data do not exclude this interaction, such an interaction does not account for the selection experiment with recombinant U1C protein. Taking into account our biochemical data (Fig. 3), another possibility is that the L13 mutation allows

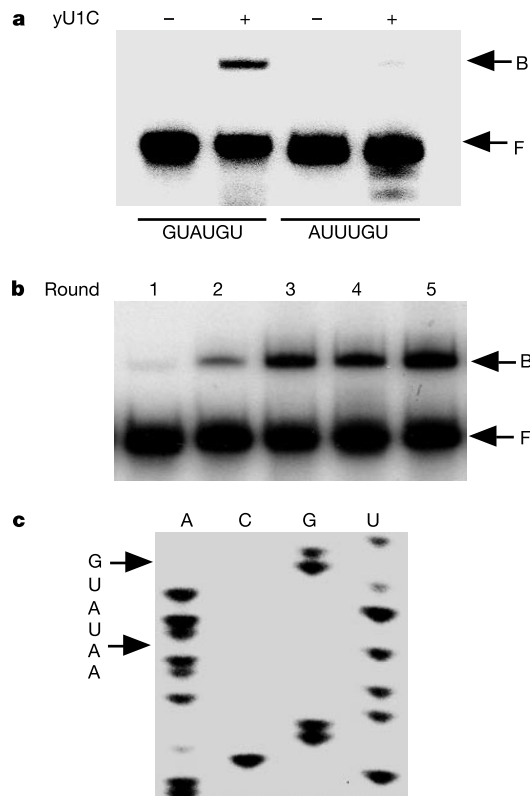


Figure 4 Characterization of sequence specificity of recombinant yU1C. **a**, EMSA for recombinant yU1C with either wild-type or 5'-splice-site mutant RNA. Bands corresponding to free (F) and bound (B) RNA are indicated on the right. **b**, EMSA of yU1C and RNA pools after successive rounds of selection, as indicated. **c**, A single winning sequence selected by yU1C was identified by DNA sequencing. Identical results were obtained with two independent preparations of U1C and starting RNA pools. The initial pools were verified as random.

spliceosome assembly to proceed without base pairing. This interpretation is based on the apparent protein-only recognition of the 5'-splice-site region by U1 snRNP at low temperature (Fig. 3). Base pairing requires more elevated temperatures, which indicates that there is a temperature-dependent step required for the U1 snRNA/5'-splice-site region interaction. In the U1C-depleted extract, however, base pairing occurs at low temperature, which suggests that the U1C protein suppresses base pairing at low temperature. This fits well with data that indicate that U1C limits pre-mRNA access to the 5' end of U1 snRNA¹⁵.

Together, these results suggest a model in which the initial interaction between U1 snRNP and the pre-mRNA 5'-splice-site region is mediated by protein. U1C contributes to this 5'-splice-site selection, which is then followed by the canonical U1 snRNA/5'-splice-site base pairing. The temperature-dependent step might reflect a rate-limiting conformational change required for base pairing. This might even be a unimolecular structural rearrangement, by analogy to ribosome assembly¹⁶. On the basis of *in vivo* evidence, we had previously considered that U1C might function as a local, dedicated RNA chaperone to accelerate duplex formation—in addition to its putative contribution to duplex stabilization¹⁵. In contrast to the mammalian system^{17–21}, there is no evidence in the yeast system for U1 snRNP-independent splicing, and there is no *in vivo* evidence for yeast 5'-splice-site recognition in the absence of U1 snRNP function.

RNA/protein interactions before RNA/RNA base pairing also form the basis of the current model for the U2 snRNP/branchpoint

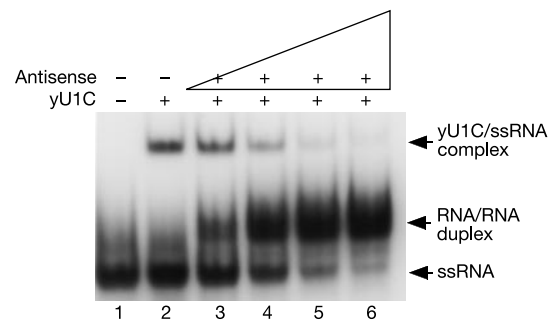


Figure 5 yU1C binding to single-stranded (ss) RNA and duplex substrates. **a**, 24-nucleotide antisense RNA (GGTCCATATTATTACCATTTTG), complementary to the GUAUAA-containing winner 72-nucleotide sequence, was used. ³²P-labelled winner 72-nucleotide was pre-incubated for 10 min with an increasing amount of cold antisense RNA under EMSA conditions, before the addition of yU1C protein (lanes 3–6). There was no antisense RNA in lane 2 and no protein in lane 1. Bands corresponding to ss RNA, duplex RNA and the yU1C/RNA complex are indicated on the right.

interaction, where proteins identify and bind to the branchpoint region before U2 snRNA/branchpoint base pairing^{22–24}. The stability of the base-paired complexes may make the snRNAs ill-suited to function as the initial selectors for pre-mRNA substrates²⁵. Although we have no data on the mammalian system, the notion that U1C acts as an initial 5'-splice-site selector fits with a study indicating that the absence of human U1C from U1 snRNP decreases the on-rate but not the off-rate of U1 snRNA/5'-splice-site base pairing²⁶. There are other proteins that probably recognize the 5' splice site in a sequence-specific manner, that is, there are direct contacts between the U5 snRNP protein Prp8 and the 5'-splice-site GU^{27,28}. There is also evidence from an *in vitro* yeast *trans*-splicing system for early, ATP-dependent contacts between U4/U6 snRNP and the 5' splice site²⁹. If U1C can identify and recruit potential 5'-splice-site sequences, then the canonical base pairing might have a proofreading function in addition to contributing to initial 5'-splice-site choice. This view suggests that 5'-splice-site selection has at least three steps comprising sequential interactions with U1C, U1 snRNA and U6 snRNA. □

Methods

yU1C overexpression and purification

We cloned the yU1C open reading frame into pETBlue-2, which incorporates a polyhistidine tag at the carboxy terminus. The recombinant protein was overexpressed in either *Escherichia coli* strain Origami (DE3)pLacI or Tuner (DE3)pLacI and purified according to the manufacturer's protocol (Novagen). We dialysed the eluate twice for 1.5 h against 1 l of buffer D (20 mM HEPES-KOH (pH 7.9), 0.2 mM EDTA, 50 mM KCl, 0.5 mM dithiothreitol and 20% glycerol).

DNA pool construction

The double-stranded DNA pool was synthesized by polymerase chain reaction (PCR). The upstream primer R1 (5'-GCGGAATTCTAATACGACTCACTATAGGTCGAGACTAGC AATAAC-3') contains an *Eco*RI site and a T7 promoter; the downstream primer R3 (5'-GCGGATCCTCAATATTACGTGTCCT-3') contains a *Bam*HI site. Oligonucleotide R2 (5'-CTCAATATTACGTGTCCTAAAGCCTCCTTAGTCCATATTANNNNNCA TTTTGTATTGTAGTCTCGACC-3'), where N represents a degenerate nucleotide, was used as a template for PCR with primers R1 and R3. The generated RNA sequence is identical to the 72-nucleotide described previously⁸, except for the six random nucleotides at the 5'-splice-site region.

EMSA

Recombinant yU1C was incubated with either radiolabelled individual RNA substrate or RNA pool in 12.5 μ l of binding buffer (25 mM Tris-HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA and 0.2 μ g of transfer RNA) for 30 min at 25 °C. We separated the reaction mixtures on 6% native polyacrylamide gels at 4 °C.

Filter binding assay

Radiolabelled RNA was incubated with recombinant yU1C in 20 μ l of binding buffer (see

above) for 30 min at 25 °C. Mixtures were passed over nitrocellulose filters under vacuum. Filters were washed with 3 ml of binding buffer before RNA recovery.

Other methods

We carried out ultraviolet crosslinking, immunoprecipitation, RNase H treatment and U1C-depletion as described^{11,15}.

Received 19 April; accepted 11 June 2002; doi:10.1038/nature00947.

- Burge, C. B., Tuschl, T. & Sharp, P. A. In *The RNA World II* (eds Gesteland, R. R., Cech, T. R. & Atkins, J. F.) 525–560 (Cold Spring Laboratory Harbor Press, Cold Spring Harbor, 1999).
- Séraphin, B. & Rosbash, M. Identification of functional U1 snRNA–pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**, 349–358 (1989).
- Michaud, S. & Reed, R. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.* **5**, 2534–2546 (1991).
- Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. & Steitz, J. A. Are snRNPs involved in splicing? *Nature* **283**, 220–224 (1980).
- Zhuang, Y. & Weiner, A. M. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**, 827–835 (1986).
- Séraphin, B., Kretzner, L. & Rosbash, M. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *EMBO J.* **7**, 2533–2538 (1988).
- Siliciano, P. G. & Guthrie, C. 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes Dev.* **2**, 1258–1267 (1988).
- Zhang, D. & Rosbash, M. Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev.* **13**, 581–592 (1999).
- Puig, O., Gottschalk, A., Fabrizio, P. & Séraphin, B. Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes Dev.* **13**, 569–580 (1999).
- Chen, J. Y. *et al.* Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol. Cell* **7**, 227–232 (2001).
- Du, H. & Rosbash, M. Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA–pre-mRNA base pairing. *RNA* **7**, 133–142 (2001).
- Lund, M. & Kjems, J. Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end. *RNA* **8**, 166–179 (2002).
- Heinrichs, V., Bach, M., Winkelmann, G. & Lührmann, R. U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science* **247**, 69–72 (1990).
- Jamison, S. F. *et al.* U1 snRNP–ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. *Nucleic Acids Res.* **23**, 3260–3267 (1995).
- Tang, J., Abovich, N., Fleming, M., Séraphin, B. & Rosbash, M. Identification and characterization of a yeast homolog of U1 snRNP-specific protein C. *EMBO J.* **16**, 4082–4091 (1997).
- Traub, P. & Nomura, M. Structure and function of *Escherichia coli* ribosomes. VI. Mechanism of assembly of 30S ribosomes studied *in vitro*. *J. Mol. Biol.* **40**, 391–413 (1969).
- Crispino, J. D., Blencowe, B. J. & Sharp, P. A. Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science* **265**, 1866–1869 (1994).
- Tarn, W. Y. & Steitz, J. A. SR proteins can compensate for the loss of U1 snRNP functions *in vitro*. *Genes Dev.* **8**, 2704–2717 (1994).
- Konforti, B. B. & Konarska, M. M. A short 5' splice site RNA oligo can participate in both steps of splicing in mammalian extracts. *RNA* **1**, 815–827 (1995).
- Crispino, J. D. & Sharp, P. A. A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes Dev.* **9**, 2314–2323 (1995).
- Crispino, J. D., Mermoud, J. E., Lamond, A. I. & Sharp, P. A. Cis-acting elements distinct from the 5' splice site promote U1-independent pre-mRNA splicing. *RNA* **2**, 664–673 (1996).
- Valcarcel, J., Gaur, R. K., Singh, R. & Green, M. R. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* **273**, 1706–1709 (1996).
- Abovich, N. & Rosbash, M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403–412 (1997).
- Berglund, J. A., Chua, K., Abovich, N., Reed, R. & Rosbash, M. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**, 781–787 (1997).
- Herschlag, D. Implications of ribozyme kinetics for targeting the cleavage of specific RNA molecules *in vivo*: more isn't always better. *Proc. Natl Acad. Sci. USA* **88**, 6921–6925 (1991).
- Rossi, F. *et al.* Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5' splice site–U1 snRNP interaction. *J. Biol. Chem.* **271**, 23985–23991 (1996).
- Reyes, J. L., Kois, P., Konforti, B. B. & Konarska, M. M. The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA* **2**, 213–225 (1996).
- Maroney, P. A., Romfo, C. M. & Nilsen, T. W. Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol. Cell* **6**, 317–328 (2000).
- Johnson, T. L. & Abelson, J. Characterization of U4 and U6 interactions with the 5' splice site using a *S. cerevisiae* *in vitro* trans-splicing system. *Genes Dev.* **15**, 1957–1970 (2001).

Acknowledgements

We thank current and ex-colleagues for discussions, and B. Séraphin and M. Moore for comments on the manuscript. H.D. was supported by a Charles A. King Trust Fellowship. The work was also supported by the National Institutes of Health.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.R. (e-mail: rosbash@brandeis.edu).

RNA aptamers as reversible antagonists of coagulation factor IXa

Christopher P. Rusconi*, Elizabeth Scardino*, Juliana Layzer*, George A. Pitoc*, Thomas L. Ortel†, Dougald Monroe‡ & Bruce A. Sullenger*

* Department of Surgery, Program in Combinatorial Therapeutics;

† Departments of Medicine and Pathology, Duke University Medical Center, Durham, North Carolina 27710, USA

‡ Center for Thrombosis and Haemostasis, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599, USA

Many therapeutic agents are associated with adverse effects in patients. Anticoagulants can engender acute complications such as significant bleeding that increases patient morbidity and mortality¹. Antidote control provides the safest means to regulate drug action. For this reason, despite its known limitations and toxicities, heparin use remains high because it is the only anticoagulant that can be controlled by an antidote, the polypeptide protamine^{2–4}. To date, no generalizable strategy for developing drug–antidote pairs has been described. We investigated whether drug–antidote pairs could be rationally designed by taking advantage of properties inherent to nucleic acids to make antidote-controlled anticoagulant agents. Here we show that protein-binding oligonucleotides (aptamers) against coagulation factor IXa are potent anticoagulants. We also show that oligonucleotides complementary to these aptamers can act as antidotes capable of efficiently reversing the activity of these new anticoagulants in plasma from healthy volunteers and from patients who cannot tolerate heparin⁵. This generalizable strategy for rationally designing a drug–antidote pair thus opens up the way for developing safer regulatable therapeutics.

To determine if properties inherent to nucleic acids can be used to develop an antidote-controlled anticoagulant, we first sought to generate an aptamer with anticoagulant activity. Procoagulant proteins that promote fibrin clot formation have been targeted in the development of many anticoagulant agents⁶, and anticoagulant aptamers have been isolated against coagulation factors VIIa⁷ and thrombin^{8–10}. Here we describe the isolation of aptamers specific for coagulation factor IXa (FIXa).

We employed iterative *in vitro* selection techniques^{11,12} to screen a nucleic-acid-based combinatorial library containing about 10¹⁴ species for those members capable of binding FIXa with high affinity. To ensure that the resultant aptamers would be stable in human plasma, the starting library contained 2'-fluoropyrimidines¹³. *In vitro* selection was performed for eight rounds against FIXa. The RNAs present in the round-eight library were converted to complementary DNAs and sequenced. Sixteen of the RNAs obtained bind FIXa, and all share a conserved primary sequence and secondary structure (Fig. 1a, b). Of these RNAs, aptamer 9.3 bound FIXa with the highest affinity (dissociation constant $K_d = 0.65 \pm 0.2$ nM). Covariation analysis of this sequence family¹⁴ aided in the generation of a truncated version of this aptamer, termed 9.3t ($K_d = 0.58 \pm 0.1$ nM), and an inactive mutant version, 9.3tM (K_d for FIXa > 10 μ M) (Fig. 1b). Aptamer 9.3t exhibits greater than 5,000-fold specificity for FIXa versus the structurally similar coagulation factors VIIa, Xa, XIa and activated protein C (K_d values > 5 μ M). Attachment of a polyethylene glycol of relative molecular mass $M_r = 40,000$ to the 5' end of an aptamer has been shown to enhance the bioavailability of aptamers *in vivo*^{15,16}, and its attachment to aptamer 9.3t had a nominal impact on the affinity of this aptamer for FIXa, (K_d of Peg-9.3t, 2.83 ± 0.4 nM).

We next determined if aptamer 9.3t inhibits FIXa activity. *In vivo*, FIXa forms a complex with coagulation factor VIIIa on a cell surface

and catalyses the cleavage of FX¹⁷. Therefore, we pre-assembled the FIXa–FVIIIa enzyme complex on a liposome surface, and measured the ability of aptamer 9.3t to inhibit the activation of FX by this complex (Fig. 2a). Aptamer 9.3t, but not 9.3tM, completely blocked FX cleavage by the FIXa–FVIIIa enzyme complex. Next, we determined if aptamer 9.3t could block the activity of FIXa in human plasma (Fig. 2b). Specific inhibitors of FIXa prolong the clotting time of plasma as measured in activated partial thromboplastin time (APTT) assays, but do not prolong the clotting time of plasma when measured in prothrombin time (PT) assays¹⁷. Aptamers 9.3t and Peg-9.3t prolonged the APTT clotting times in a dose-dependent manner, but had no effect on clotting times measured in PT assays, demonstrating that the aptamers specifically inhibit the FIXa pathway in plasma. Both aptamers were capable of increasing the APTT clotting time to the level observed in plasma from individuals deficient in FIX. By comparison, control aptamer 9.3tM had no effect in either clotting assay.

FIXa has been described as a ‘safe’ target for anticoagulation and antithrombotic therapy because anti-FIXa agents have exhibited reduced bleeding risks in animal models when used at their minimal effective doses as compared to heparin^{18–20}. However, at marginally higher doses, the same agents have exhibited bleeding profiles no different from that of heparin^{18–20}. Such an experience in well-controlled animal studies suggests that in the clinical setting, tight regulation of FIXa inhibitors would enhance their safety and facilitate their medical use. Antidote control is preferable to other methods of drug regulation such as pharmacokinetic strategies, as patient pathophysiology can significantly affect the rate of drug clearance. Therefore, we sought to design an antidote that could tightly regulate the anticoagulant activity of aptamer 9.3t. We hypothesized that by altering the shape of an aptamer from an active to an inactive conformation, we could modulate the target binding and inhibitory activity of the aptamer. Moreover, we speculated that any agent that critically alters the conformation of an aptamer could act as an antidote for that aptamer.

To test this hypothesis, we investigated the use of oligonucleotides complementary to aptamer 9.3t as potential antidotes for this anticoagulant (Fig. 3a). We designed a series of antidote oligonucleotides complementary to different portions of aptamer 9.3t (Fig. 3b). To compare the antidote activity of these oligonucleotides, we evaluated them under clinically relevant conditions. First, we anticoagulated human plasma with aptamer 9.3t. Then we added increasing amounts of antidote oligonucleotides to the anticoagulated plasma and measured the plasma’s ability to clot 10 min later (Fig. 3c). We observed a wide range in the ability of the various antidote oligonucleotides to reverse the anticoagulant activity of aptamer 9.3t, with antidote 5-2 being the most effective. The relative

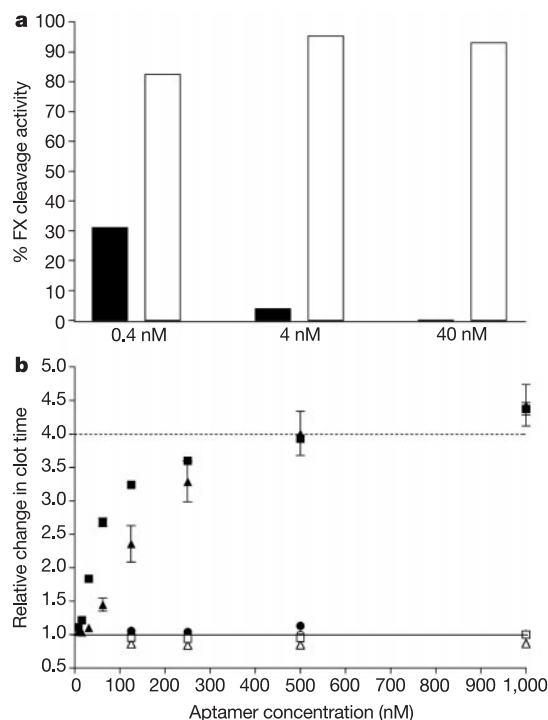


Figure 2 Inhibition of FIXa function by aptamers 9.3t and Peg-9.3t. **a**, The rate of FX activation by the FVIII–FIXa enzyme complex was measured in the presence of varying concentrations of 9.3t (filled bars) or 9.3tM (open bars). **b**, Varying concentrations of aptamers 9.3t, Peg-9.3t or 9.3tM were added to normal human plasma and clotting times were measured in APTT or PT assays: 9.3t APTT (filled squares), PT (open squares); Peg-9.3t APTT (filled triangles), PT (open triangles); 9.3tM APTT (filled circles), PT (open circles). The dashed line indicates the observed increase in the APTT of human plasma deficient in FIX (less than 1% FIX activity) as compared with normal human plasma (solid line). Data shown is the mean $\pm$ s.e.m. for three independent measurements.

ability of these antidote oligonucleotides to bind aptamer 9.3t directly correlates with their effectiveness as antidotes in human plasma (Fig. 3d).

An antidote must be able to rapidly reverse an anticoagulant’s activity, and dosing of an antidote would be simplified if the antidote’s effects were durable. To determine if antidote 5-2 met these criteria, we further characterized its activity against aptamer Peg-9.3t, as the pegylated form of the aptamer is more likely to be used in the clinic. Antidote 5-2 completely reverses the anticoagulant activity of aptamer Peg-9.3t in 10 min and does so at

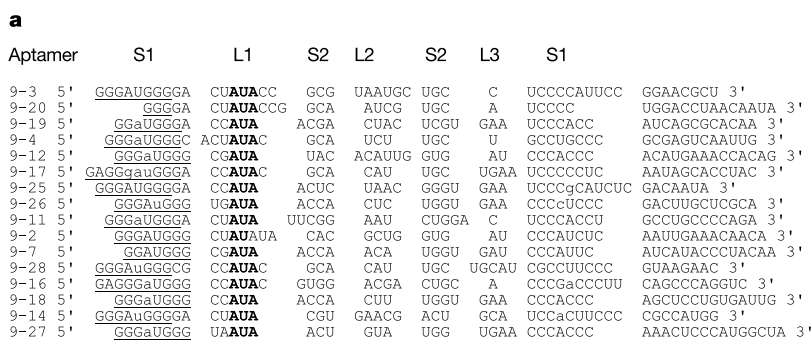


Figure 1 Sequences and secondary structure of aptamers isolated against FIXa.

a, Sequences are aligned on the basis of their predicted secondary structure and listed in order of decreasing affinity for FIXa. Shown are sequences that originated from the randomized region of the library used in the SELEX process and portions of the 5’ fixed region sequence proposed to participate in the conserved secondary structure (fixed-

region sequences are underlined, conserved loop sequence is in bold) stem (S), loop (L). For some of the sequences, mis-pairings are present in stem 1 (S1), and are denoted by lower-case letters. **b**, Predicted secondary structure of aptamer 9.3t, a 35-nucleotide truncate of aptamer 9.3 (idT, inverted deoxythymidine). A double-point mutation that yields an inactive aptamer, 9.3tM, is also shown.

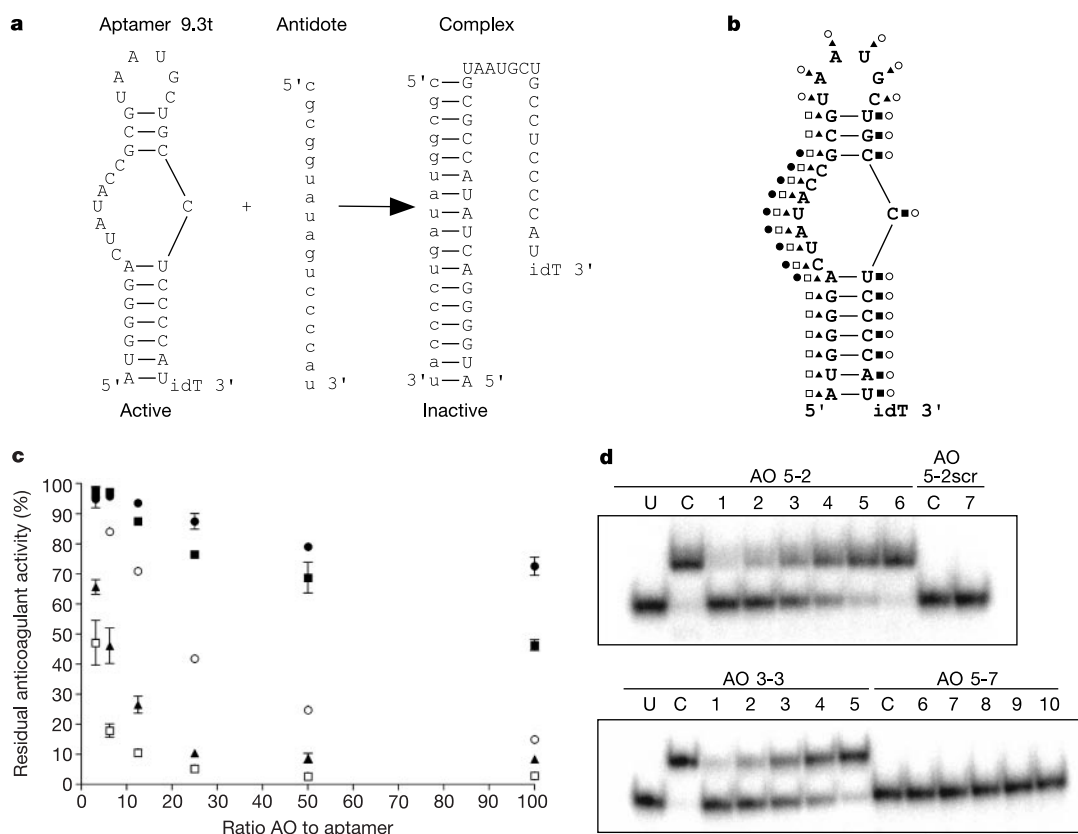


Figure 3 Antidote oligonucleotides to aptamer 9.3t. **a**, Cartoon depicting an antidote oligonucleotide binding to aptamer 9.3t to form an inactive aptamer–antidote complex. **b**, Antidote oligonucleotides used in this study: 5-1 (filled triangles), 5-2 (open squares), 5-7 (filled circles), 3-1 (open circles), 3-3 (filled squares). Placement of the symbols depicts the positions of complementarity between the antidote and the aptamer. All antidote oligonucleotides are 2'-O-methyl modified. **c**, Reversal of anticoagulant activity of aptamer 9.3t (50 nM) by antidote oligonucleotides (AO) in human plasma. Data are the

average $\pm$ range for duplicate measurements. **d**, Native gel analysis of antidote oligonucleotide binding to aptamer 9.3t. Untreated aptamer (U), aptamer denatured and complexed with antidote before gel loading (C). The upper panel shows the antidote molar ratio to aptamer increasing from 0.5:1 to 16:1 in twofold increments (lanes 1–6) and the 16:1 ratio of scrambled control antidote to aptamer (lane 7). The lower panel shows the antidotes molar ratio to aptamer increasing from 1:1 to 16:1 in twofold increments (lanes 1–5 and 6–10).

slightly lower concentrations than required for aptamer 9.3t (Fig. 4a). By comparison, a control 'scrambled' version of antidote 5-2 (5-2scr) did not affect the anticoagulation activity of the aptamers or the clotting activity of the plasma. To evaluate the kinetics of onset and persistence of antidote action, we measured the time required for antidote 5-2 to reverse the anticoagulant activity of aptamer Peg-9.3t and the durability of this reversal once established. The antidote activity of oligonucleotide 5-2 towards aptamer Peg-9.3t is rapid, and concentration dependent in human plasma (Fig. 4b). Finally, after its addition, antidote 5-2 reversed the aptamer's activity for over 5 h, indicating that the activity of this antidote is effectively irreversible in human plasma (Fig. 4c). In addition, plasma treated with aptamer Peg-9.3t but given no antidote remained anticoagulated over the time course of this experiment, demonstrating that aptamer Peg-9.3t is quite stable in human plasma, as expected for an RNA containing 2'-fluoropyrimidine modifications and an inverted deoxythymidine on its 3' end²¹.

The ability to control the anticoagulant activity of heparin with protamine enables safer treatment of patients undergoing procedures requiring a high level of anticoagulation, in whom the post-procedural risk of haemorrhage is high. However, approximately 3–5% of the estimated 12 million patients who receive heparin each year develop a drug-induced immunologic response, termed heparin-induced thrombocytopenia (HIT)²², that contraindicates further treatment of these patients with heparin⁵. Alternative anticoagulants are available for treatment of these patients, but haemorrhagic complications and recurrent life- and limb-threaten-

ing thromboembolism are common during treatment^{23,24}. Moreover, none of these anticoagulants can be neutralized by a reversing agent, which limits physician control during treatment of these patients. Therefore, we investigated the ability of aptamer Peg-9.3t and antidote 5-2 to serve as an anticoagulant–antidote pair in plasma samples from six patients with HIT, three with end-stage renal disease requiring haemodialysis and thus repeated anticoagulation, and three with thromboembolic complications requiring anticoagulant therapy. Aptamer Peg-9.3t prolonged the APTT clotting times of plasma from all six patients, and antidote 5-2 was able effectively to reverse this anticoagulant activity to the pre-treatment baseline of each patient (Fig. 5). Importantly, two patients were receiving anticoagulant therapy at the time samples were taken (patient 3 on danaparoid sodium and patient 6 on warfarin), demonstrating that in patient plasma, the aptamer–antidote pair can function independently of an 'on board' anticoagulant.

We have demonstrated that drug–antidote pairs can be rationally designed by taking advantage of properties inherent to nucleic acids to make an antidote-controlled anticoagulant aptamer. To achieve this end, we adopted an approach employing complementary antidote oligonucleotides, as this approach is straightforward and generalizable to any aptamer. We have generated an antidote oligonucleotide that effectively, rapidly and durably inactivates a new anticoagulant aptamer in human plasma under parameters set as the clinical standard by heparin and protamine³. Moreover, the demonstration that aptamer Peg-9.3t and its antidote function in

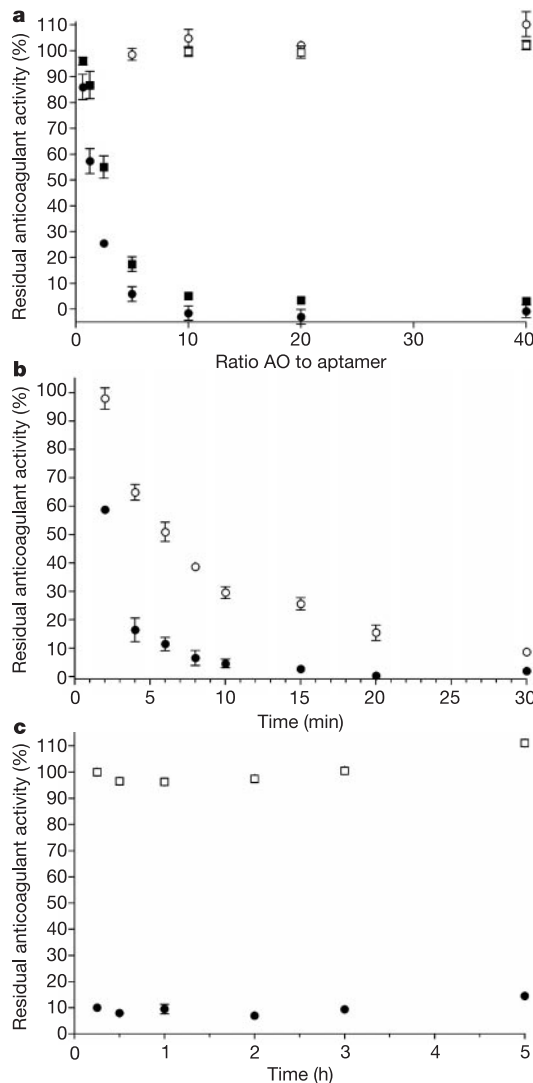


Figure 4 Properties of antidote oligonucleotide 5-2. **a**, The ability of antidote oligonucleotides 5-2 or 5-2scr to reverse the anticoagulant activity of aptamers 9.3t and Peg-9.3t (aptamers at 125 nM) was measured in pooled human plasma: 9.3t + 5-2 (filled squares); 9.3t + 5-2scr (open squares); Peg-9.3t + 5-2 (filled circles); and Peg-9.3t + 5-2scr (open circles). Data is mean $\pm$ s.e.m. for three independent measurements. **b**, Kinetics of the reversal of the anticoagulant activity of aptamer Peg-9.3t by antidote oligonucleotide 5-2 in human plasma, Peg-9.3t + 2 $\times$ 5-2 (open circles), +10 $\times$ 5-2 (filled circles). Data is average $\pm$ range for two independent measurements. **c**, The duration of the reversal of the anticoagulant activity of aptamer Peg-9.3t by antidote oligonucleotide 5-2 was measured in human plasma (filled circles). The persistence of the anticoagulant activity of Peg-9.3t without the antidote in human plasma is shown for comparison (open squares). Data is the average $\pm$ the range for two independent measurements.

plasma from patients suffering from HIT is a first step towards providing such an anticoagulant–antidote pair for this patient population. The relative excess of antidote over aptamer necessary to achieve these results is compatible with the use of this approach *in vivo*, and current efforts are aimed at translating these results obtained with patient samples into animal models. Over the past decade much progress has been made in translating aptamers into therapeutics, including the demonstration of efficacy of several aptamers in animal models, and phase III testing of an aptamer against vascular endothelial growth factor in clinical trials^{21,25}. However, the rationale for therapeutic aptamer development has been questioned because other classes of antagonists, such as

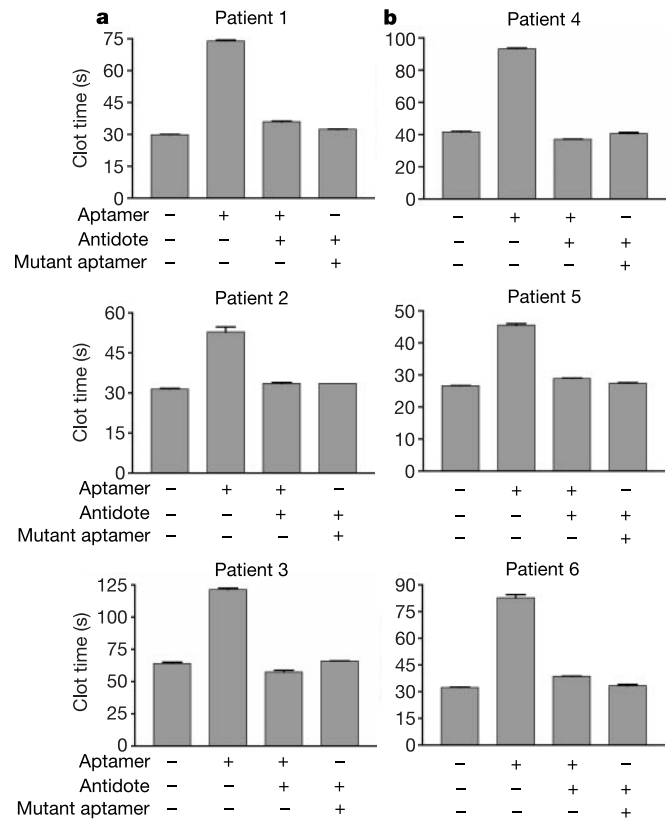


Figure 5 Antidote-controlled anticoagulation of plasma from patients with HIT. **a, b**, The activity of aptamer Peg-9.3t and antidote 5-2 were tested in plasma from haemodialysis-dependent patients diagnosed with HIT (**a**) and in plasma from patients suffering from thromboembolic complications of HIT (**b**). Plasma samples were treated as indicated: aptamer, 125 nM Peg-9.3t; antidote, 1.25 μ M AO 5-2; and mutant aptamer, 125 nM 9.3tM. Data is reported in seconds (s) and is the average $\pm$ range of duplicate measurements.

humanized monoclonal antibodies, have already proved useful as therapeutic agents²⁶. The ability to rationally design antidotes to alter the shape and activity of aptamers is unique to aptamers. We believe this property distinguishes aptamers from other classes of antagonists, such as antibodies, peptides or small molecules. We recommend the clinical development of safer regulatable therapeutics based upon aptamer–antidote pairs for indications in which the acute side effects of treatment may lead to increased morbidity and mortality. □

Methods

Generation of aptamers

In vitro selection procedures were carried out as previously described^{7,10,27}. The sequence of the starting RNA combinatorial library was 5'-GGGAGAGAGGAAGAGGGAUGGG-N₄₀-CAUAAACCCAGAGGUCGAU-3', where N₄₀ represents 40 contiguous nucleotides with equimolar A, G, C and U at each position. 2' F cytidine triphosphate and uridine triphosphate were purchased from TriLink Biotechnologies and incorporated into RNA libraries by *in vitro* transcription as described¹⁰. The selection was carried out in selection buffer (20 mM HEPES, pH 7.4, 50 mM NaCl, 2 mM CaCl₂, and 0.01% bovine serum albumin (BSA) at 37 °C, using filtration through nitrocellulose membranes to separate FIXa-RNA complexes from unbound RNA.

Binding studies

K_d values were determined using double-filter nitrocellulose filter binding assays⁷. All binding studies were performed in physiologic binding buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 2 mM CaCl₂, and 0.01% BSA) at 37 °C, and all coagulation factors were purchased from Haematologic Technologies Inc. Aptamer 9.3t, 9.3tM and Peg-9.3t were synthesized and purified by Dharmacon Research Inc. K_d values of aptamers 9.3 and 9.3t binding to FIXa and other coagulation factors were determined in direct binding assays, and K_d values for aptamers Peg-9.3t and 9.3tM binding to FIXa were determined in competition binding assays. Competition binding studies were performed and results

analysed as described¹⁰, with a final FIXa concentration of 1 nM. All data presented is the mean $\pm$ s.e.m. for at least three independent experiments.

Enzyme assays

Factor VIII (Hemophil M, Baxter Healthcare) was activated with thrombin, then the FVIIIa solution was added to factor IXa and phospholipid vesicles (lipids purchased from Avanti Polar Lipids and vesicles prepared as in ref. 28) and allowed to form the FIXa–FVIIIa enzyme complex. Factor X and aptamer were then added, and factor X activation measured with a chromogenic substrate (Pefachrome fXa, Centerchem Inc.). Final concentrations were: factor IXa, 0.2 nM; factor VIIIa, 0.7 nM; lipid, 40 μ M; factor X, 135 nM; and factor Xa substrate, 1 mM. The rate of factor Xa generation was determined by a second-order fit to the data. The %FX cleavage activity is 100 times the rate of FXa generation in the presence of aptamer divided by rate in the absence of aptamer.

Clotting assays

Activated partial thromboplastin time (APTT) assays were performed using a model ST4 mechanical coagulometer (Diagnostic Stago Inc.). Aptamer in binding buffer (5 μ l) without BSA or binding buffer without BSA alone was added to pooled normal human plasma (50 μ l) (George King Biomedical), and incubated for 5 min at 37 °C. MDA platelin (50 μ l) (bioMerieux) was then added and allowed to activate the plasma for 5 min, followed by the addition of 25 mM CaCl₂ (50 μ l) to initiate the clotting reaction. Data is expressed as the relative change in clot time; the clot time in the presence of aptamer divided by the clot time in the presence of buffer alone. All reactions were performed in duplicate, and only duplicates differing by <10% were used in analysis.

Prothrombin time (PT) clotting assays were performed as previously described⁷ except that 5 μ l of aptamer was added to 50 μ l of normal pooled human plasma, and all reactions were carried out at 37 °C.

Antidote assays

Antidote oligonucleotides were synthesized and purified by Dharmacon Research, Inc. Antidote activity was measured 10 min after antidote addition to plasma containing aptamer in APTT clotting assays. Briefly, human plasma was anticoagulated with aptamer, antidote oligonucleotide (5 μ l) was added and the incubation continued for 5 min before the addition of MDA platelin. Antidote activity is expressed as the per cent residual anticoagulant activity *T* of the aptamer, which is:

$$[1 - (T_{\text{aptamer alone}} - T_{\text{aptamer+antidote}}) / (T_{\text{aptamer alone}} - T_{\text{baseline}})] \times 100.$$

For measuring the kinetics of the onset of antidote activity, the incubation time of the plasma following MDA platelin addition was reduced to 1 min to allow for shorter timepoints to be measured. This increased the baseline APTT from 30.2–32.5 s to 34.2–36.8 s.

Gel shift assays were performed essentially as described²⁹. Aptamer 9.3t (50 nM with trace ³²P-labelled) was incubated for 10 min with varying concentrations of antidote oligonucleotide in binding buffer without BSA before loading on a 12% polyacrylamide gel containing 2 mM CaCl₂. Gels were run for 3 h at 300 V and visualized using a Storm 840 Phosphorimager (Molecular Dynamics).

Patient samples

Plasma samples from six patients with HIT were studied. Clinical criteria for the diagnosis of HIT included thrombocytopenia and/or new or recurrent thrombosis after five or more days of heparin therapy³. Serologic criteria included a positive heparin-induced platelet aggregation assay and/or elevated heparin/platelet factor 4 antibody levels detected by enzyme-linked immunosorbent assay (ELISA) (GTI Inc.). Five patients met both clinical and serologic criteria; one patient fulfilled clinical criteria but had negative serologic studies. The Institutional Review Board at Duke University Medical Center approved these studies, and informed consent was obtained from all patients.

Received 12 February; accepted 18 June 2002; doi:10.1038/nature01058.

- Levine, M. N., Raskob, G., Landefeld, S. & Kearon, C. Hemorrhagic complications of anticoagulant treatment. *Chest* **119**, 1085–1215 (2001).
- Hirsch, J., Anand, S. S., Halperin, J. L. & Fuster, V. Guide to anticoagulant therapy: Heparin: A statement for healthcare professionals from the American Heart Association. *Circulation* **103**, 2994–3018 (2001).
- Carr, J. A. & Silverman, N. The heparin-protamine interaction. A review. *J. Cardiovasc. Surg.* **40**, 659–666 (1999).
- Pifarre, R., Walenga, J. M. & Fareed, J. In *New Anticoagulants for the Cardiovascular Patient* (ed. Pifarre, R.) 1–7 (Hanley and Belfus, Philadelphia, 1997).
- Warkentin, T. E., Chong, B. H. & Greinacher, A. Heparin-induced thrombocytopenia: towards consensus. *Thromb. Haemost.* **79**, 1–7 (1998).
- Johnson, K. & Hung, D. Novel anticoagulants based on inhibition of the factor VIIa/tissue factor pathway. *Coron. Artery Dis.* **9**, 83–87 (1998).
- Rusconi, C. P., Yeh, A., Lyster, H. K., Lawson, J. H. & Sullenger, B. A. Blocking the initiation of coagulation by RNA aptamers to factor VIIa. *Thromb. Haemost.* **84**, 841–848 (2000).
- Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566 (1992).
- Tasset, D. M., Kubik, M. F. & Steiner, W. Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J. Mol. Biol.* **272**, 688–698 (1997).
- White, R. *et al.* Generation of species cross-reactive aptamers using “toggle” SELEX. *Mol. Ther.* **4**, 567–573 (2001).
- Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
- Ellington, A. D. & Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).

- Pieken, W. A., Olsen, D. B., Benseler, E., Aupur, H. & Eckstein, F. Kinetic characterization of ribonuclease-resistant 2'-modified hammerhead ribozymes. *Science* **253**, 314–317 (1991).
- Davis, J. P., Janjic, N., Javornik, B. E. & Zichi, D. A. Identifying consensus patterns and secondary structure in SELEX sequence sets. *Methods Enzymol.* **267**, 302–314 (1996).
- Watson, S. R. *et al.* Anti-L-selectin aptamers: binding characteristics, pharmacokinetic parameters, and activity against an intravascular target in vivo. *Antisense Nucleic Acid Drug Dev.* **10**, 63–75 (2000).
- Tucker, C. E. *et al.* Detection and plasma pharmacokinetics of an anti-vascular endothelial growth factor oligonucleotide-aptamer (NX1838) in rhesus monkeys. *J. Chromatogr. B Biomed. Sci. Appl.* **732**, 203–212 (1999).
- High, K. A. & Roberts, H. R. in *Molecular Basis of Thrombosis and Hemostasis* (eds High, K. A. & Roberts, H. R.) 215–237 (Marcel Dekker, New York, 1995).
- Spanier, T. B. *et al.* Selective anticoagulation with active site-blocked factor IXA suggests separate roles for intrinsic and extrinsic coagulation pathways in cardiopulmonary bypass. *J. Thorac. Cardiovasc. Surg.* **116**, 860–869 (1998).
- Feuerstein, G. Z. *et al.* An inhibitory anti-factor IX antibody effectively reduces thrombus formation in a rat model of venous thrombosis. *Thromb. Haemost.* **82**, 1443–1445 (1999).
- Choudhri, T. F. *et al.* Targeted inhibition of intrinsic coagulation limits cerebral injury in stroke without increasing intracerebral hemorrhage. *J. Exp. Med.* **190**, 91–99 (1999).
- White, R. R., Sullenger, B. A. & Rusconi, C. P. Developing aptamers into therapeutics. *J. Clin. Invest.* **106**, 929–934 (2000).
- Campbell, K. R. *et al.* Bivalirudin in patients with heparin-induced thrombocytopenia undergoing percutaneous coronary intervention. *J. Invasive Cardiol.* **12**, 14F–19F (2000).
- Greinacher, A. *et al.* Recombinant hirudin (lepirudin) provides safe and effective anticoagulation in patients with heparin-induced thrombocytopenia: a prospective study. *Circulation* **99**, 73–80 (1999).
- Lewis, B. E. *et al.* Argatroban anticoagulant therapy in patients with heparin-induced thrombocytopenia. *Circulation* **103**, 1838–1843 (2001).
- Hicke, B. J. & Stephens, A. W. Escort aptamers: a delivery service for diagnosis and therapy. *J. Clin. Invest.* **106**, 923–928 (2000).
- van Dijk, M. A. & van de Winkel, J. G. Human antibodies as next generation therapeutics. *Curr. Opin. Chem. Biol.* **5**, 368–374 (2001).
- Fitzwater, T. & Polisky, B. A SELEX primer. *Methods Enzymol.* **267**, 275–301 (1996).
- Hope, M. J., Bally, M. B., Webb, G. & Cullis, P. R. Production of large unilamellar vesicles by a rapid extrusion procedure. Characterization of size distribution, trapped volume and ability to maintain a membrane potential. *Biochim. Biophys. Acta* **812**, 55–65 (1985).
- Silverman, S. K. & Cech, T. R. Energetics and cooperativity of tertiary hydrogen bonds in RNA structure. *Biochemistry* **38**, 8691–8702 (1999).

Acknowledgements

We thank R. Califf and R. Harrington for their insight into the need for antidote-controlled anticoagulant and antithrombotic agents. This work was supported by grants from the American Heart Association to C.P.R. and the National Institutes of Health to B.A.S., T.L.O. and D.M.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on *Nature's* website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to C.P.R. (e-mail: c.rusconi@cgct.duke.edu) or B.A.S. (e-mail: b.sullenger@cgct.duke.edu).

corrigendum

Sub-ångstrom resolution using aberration corrected electron optics

P. E. Batson, N. Dellby & O. L. Krivanek

Nature **418**, 617–620 (2002).

For this Letter the disclosure form for the declaration of competing financial interests was incorrectly filled out because of a misunderstanding. The statement should have read: ‘The authors declare competing financial interests: details accompany the paper on *Nature's* website (<http://www.nature.com/nature>).’ The details on the website should have read: ‘O.L.K. and N.D. have a personal financial interest in Nion, R&D.’ We (the authors) had no intention of misrepresenting the origin of the work. □

Putting the C into PCR

A round-up of thermocyclers and other PCR-related kit.

Replication System and TSP

Nunc

www.nuncbrand.com

Steel yourself

With pins made of stainless steel, the Nunc Replication System is capable of transferring small inocula (1.0 µl from liquid, 0.1 µl from solid supports), resulting in clear, compact spots. In applications where a slightly larger spot is an advantage, Nunc recommends using the transferable solid phase (TSP), which has a volume of 1.0 µl, producing a less dense inoculum. The system can be converted between 96- and 384-well formats using copying frame accessories. Applications include making libraries, where it can be used in the setting up and analysis stages when transferring samples to membrane. In addition, the TSP can be used in assessment of cultures in a 96-well format without duplicating or contaminating the cultures and allowing growth to proceed undisturbed by the assay. The TSP is available in three versions, PolySorp, MaxiSorp and a hybridoma screening system.



Pin up: Nunc's latest replication system.

Cells-to-cDNA II kit

Ambion

www.ambion.com

Cut to the chase

Ambion's Cells-to-cDNA produces cDNA from cultured cells in less than 2 hours with no requirement for RNA isolation. It can accommodate from 1 to 12,500 cells per reverse transcription (RT) reaction and is compatible with tissue culture cells, laser capture microdissection samples and formalin-fixed cells. The kit is suitable for gel-based or real-time RT-PCR assays and can be used with one- or two-step RT-PCR protocols that can be adapted to an automated 96-well format.

Purifier PCR enclosure

Labconco

www.labconco.com

Clear the air

Labconco's Purifier PCR enclosure has an integral blower and a 99.9% efficient HEPA filter that provides constant circulation of filtered, Class 100 air across the work area, providing a particulate-free working space. Available in 2- and 3-foot widths, the Purifier features a self-contained ultraviolet lamp with a timer that provides a five-minute exposure to deactivate DNA and RNA contaminants. Front-mounted switches for fluorescent and ultraviolet lights and the blower are located within easy reach of the operator. Side panels made of safety glass permit additional illumination of the work surface.

PCR-Cooler

Eppendorf

www.eppendorfsi.com

One for the road

The Eppendorf PCR-Cooler is a portable cooling block that maintains samples at 0 °C for over an hour and changes colour when temperatures exceed 7 °C. The compact unit is designed for cooling 0.2 ml or 0.5 ml tubes at temperatures of between 0 °C and 7 °C for approximately two hours, at an ambient temperature of 22 °C. It can be used to store and transport samples and to safely stop reactions. Tubes can be inserted individually, in a rack, in a 96-well PCR plate or in tube strips. The cooler is available in two colours for easy identification when using more than one unit.

TGradient

Biometra

www.biometra.com

Tip-top technology

The TGradient thermocycler from Biometra can test 12 annealing temperatures in a single run, helping to speed up the optimization of new primers. The instrument is fitted with an adjustable heated lid with automatic pressure control (smart lid technology), exchangeable block modules and a temperature gradient of up to 40 °C. Other features include a gold-plated silver block, a large



Hot shot: TGradient tests 12 temperatures per run.

display, low power consumption and noise emission, and uniform heating and cooling rates.

DNA labels

Cambio

www.cambio.co.uk

Beyond the pale

Fluorescent tag dyes are increasingly being used as an alternative to radiolabels in DNA probes and primers. Redmond Red and

Yakima Yellow can be incorporated into synthetic DNA molecules to allow probe detection on a variety of platforms. Eclipse Quencher is a non-fluorescent quencher that allows DNA detection probes to be used for real-time PCR applications such as measuring gene expression or detecting single nucleotide polymorphisms. Probes require efficient quenching until hybridized to the target. The fluorophore released from the target and separated from the quencher is highly fluorescent and ready for detection.

QuantiTect SYBR Green

Qiagen www.qiagen.com

Cast a discriminating dye

First items in Qiagen's new QuantiTect product line are two SYBR Green kits, one for PCR and two-step RT-PCR, and another for one-step RT-PCR. Both kits contain the fluorescent dye SYBR Green I in a master-mix format. The kits can be used on most types of real-time PCR cyclers, including ABI sequence detection systems, LightCycler and iCycler. HotStarTaq DNA polymerase in both kits prevents the formation of primer-dimers and nonspecific amplification products during reaction set-up and the first denaturation step. This eliminates inaccurate quantification due to low PCR specificity — a problem often associated with SYBR Green, which binds to any double-stranded DNA molecules present, including non-specific products and primer-dimers.

MBS[®] thermocycler

ThermoHybaid www.thermohyбайд.com

Stacks of fun

ThermoHybaid's Robot Compatible Multi-Block System (MBS[®]) for high-throughput thermal cycling has an automatic heated lid that is suitable for robotic, oil-free applications. The entire block is ejected from the front of the unit, and the system is stackable, enabling users to build a wall of thermal cycling units. The blocks reach into hooded pipetting stations for sterility, and PCR runs are automatically logged and recorded. The MBS[®] model can be integrated into existing robotic systems.

NucleoFast 96 PCR

Macherey-Nagel www.macherey-nagel.com

Clean PCR products in 20 minutes

The NucleoFast 96 PCR system makes use of ultrafiltration technology for automated and manual purification of PCR products. It facilitates the rapid clean-up of PCR products in the 96-well format and can be used manually under vacuum or centrifugation. NucleoFast 96 PCR is easily automated on standard liquid-handling instruments. The optimized ultrafiltration membrane has been developed to achieve a high recovery even for small PCR fragments (>150 bp).

Average recovery rates are 50–95% for PCR products 150–10,000 bp in size. PCR products are collected on the ultrafiltration membrane while contaminants (primers, dNTPs, salts) are filtered to waste. NucleoFast 96 PCR is available as plates only or as complete kits for more convenience.

ArchaeMaxx

Stratagene www.stratagene.com

dUMP your inhibitions

Stratagene has announced the isolation of the ArchaeMaxx polymerase-enhancing factor from *Pyrococcus furiosus* and added it to their high-fidelity enzymes and Pfu-based enzyme blends, where it eliminates a PCR inhibitor and promotes shorter extension times, higher yields and greater target length amplification. ArchaeMaxx overcomes 'dUTP poisoning', which inhibits Pfu DNA polymerase and many other archaeal proofreading polymerases, by converting poisonous dUTP to harmless dUMP and inorganic pyrophosphate.

Px2 personal thermal cycler

Thermo Life Sciences www.thermols.com

Its lids are sealed

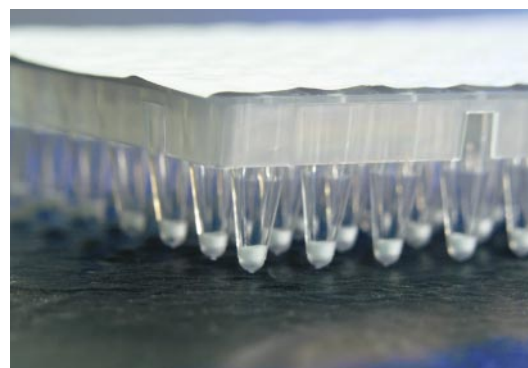
Thermo Life Sciences describes its Px2 as a user-friendly personal thermal cycler, with a high-definition user interface and associated software and navigation keypad to simplify programming and running protocols. A new heated lid design provides increased temperature and pressure and minimizes the risk of sample evaporation during the PCR run.

UltraClean

Mo Bio www.mobio.com

Take it for a spin

Mo Bio's UltraClean standard mini plasmid prep kit follows an easy spin filter kit protocol and all reagents are ready to use. It uses standard lysis reagent volumes of 250 or 350 µl and yields up to 20 mg of automated, sequence-quality, plasmid DNA. Typical A₂₆₀/A₂₈₀ ratios are 1.8–2.0.



Well done: Sprint Advantage by BD Biosciences.

Sprint Advantage

BD Biosciences www.bdbiosciences.com

Have it on a plate

Each well of the BD Sprint Advantage 96-well plate contains a complete lyophilized master mix of titanium Taq, BD TaqStart antibody, a proofreading enzyme for increased fidelity, dNTPs and an optimized PCR buffer. To use, the mixture is dissolved using diluted primers and DNA template; it is then ready to go directly for PCR.

Retrotools cDNA

Biotoools B&M Labs www.biotoools.net

An alternative for robust cDNA synthesis

The Retrotools cDNA/DNA Polymerase Kit is designed for easy reverse transcription of RNA followed by amplification of the resulting cDNA in a single step. A single enzyme performs the whole RT-PCR procedure without the need to combine viral reverse transcriptases with bacterial polymerases, rendering higher yields and overcoming secondary structures as reverse transcription is performed at high temperatures. The kit is available in liquid or gel formats.

These notes are compiled in the Nature office from information provided by the manufacturers. Press releases to be considered for publication can be sent to newproducts@nature.com.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakakatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.johnson@naturejp.com

Asia-Pacific Advertising Manager:

Keyn Johnson

naturejobs

Setting salaries straight

Science is often portrayed as an altruistic enterprise, with its participants willingly sacrificing income, personal life and job security during their long educational and training period. Although these sacrifices may be real enough, the willingness with which they are undertaken may be more of a myth.

To combat the problem, charitable organizations in the United States (see *Naturejobs* 5; 10 January 2002) and Britain have been trying to boost stipends and salaries through the funds they offer, in the hope that governments will follow suit. But the overall approach has had mixed results and unintended consequences.

On one hand, the people who receive higher stipends from foundations are understandably happy. But governments have been slow to pursue pay parity, although both the US and the UK governments are showing signs that they are at least considering slowly bumping up stipends.

But as long as a lag between the two persists, problems will inevitably occur. Mary Phillips, a programme officer at the London-based biomedical charity the Wellcome Trust, notes that, in some cases, salaries funded by the charity can be 30% higher than those for similar positions paid for by the government. Although people are not supposed to talk about salaries, scientists still know which organization is funding whom — and, indirectly, can put together a picture of pay disparity. “It can lead to some tensions in the lab,” Phillips says.

It is human nature for some resentment to occur in such situations — even in a supposedly altruistic profession. But that resentment should be channelled and directed at the government funding agencies, not the foundations who push for better scientific wages or at the people who receive them.

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS AND STUDENTS

Making the move into industry

p5

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Working your way into industry

Charles Adkins:

In academia, you have your project and it's your idea, your baby. Coming to industry, you have to work as part of a team.



What is the best way to get into the private sector? For most postdocs, it has been to apply directly for an advertised job. For a few others, undergraduate internships, summer work-study and even some postdoctorates have led to permanent jobs. But whether these forms of entry are a good way to land a long-term position depends on the company.

At drug firm Merck, most of the internships are for students working on a bachelor's degree, says Ruth McKernan, director of neuroscience research at the company's Terlings Park facility in Essex, UK. She feels this works better at Merck than having postdoc interns because the younger students have more realistic expectations. Postdocs are often angling for a permanent position, but at Merck "there may not be an opportunity for them", McKernan says.

Robert Young, vice-president of medicinal chemistry at Merck Frosst Canada in Montreal, says that his facility generally has a limited number of postdoc fellowships in both chemistry and biology. Although he agrees that industry postdocs are a potentially valuable experience for early-career scientists, "by no means would I call it an optimal way for them to find the best position in industry", he says.

POSTDOCS IN DEMAND

On the other hand, drug firm AstraZeneca actively seeks postdocs as permanent employees, bringing them into general programmes as well as specialized ventures such as its G-coupled protein-receptor research project. Five postdoc positions were created in Europe and North America for that programme, though not all have led to permanent jobs.

"We do encourage a portion of people to join us," notes Simon King, human-resources director for global discovery function at AstraZeneca's US headquarters in Wilmington, Delaware. "But some people find out they prefer to be back in academia." Still, he says about a third of their postdocs later take on permanent positions.

Why is the outlook on keeping interns so different from lab to lab? "I guess it depends on the individual," says King. For AstraZeneca, he says it is a way to bring in new ideas and simply "another of our diverse ways

of finding individuals to join the company".

Wolfgang Jarolimek, a research fellow at Terlings Park, joined Merck after an academic postdoc at the Baylor College of Medicine in Houston and an assistant professorship at the University of Heidelberg. He arrived there on a fellowship four years ago and took a permanent position a year later. The temporary stint gave him the chance to investigate drug discovery and see if he liked working in industry. Although this strategy worked for him, he cautions that his path is more the exception than the rule: taking a postdoc position can be a hindrance to future employment in the same company, just as taking a permanent academic position at the same institution as your last postdoc is seen as limiting.

LEARNING ABOUT THE CULTURE

Govindaswamy Panchamoorthy, a former postdoc and now a scientist in the AstraZeneca Enabling Science and Technology Biology group, also had a good foundation in academia — postdocs at Dana-Farber Cancer Institute and Brigham and Women's Hospital, both in Boston, Massachusetts. Wanting to take a different direction, he went to AstraZeneca's Wilmington site in 1997. After three years as a postdoc he took a permanent position and recently moved to the facility in Waltham, Massachusetts, where he develops novel drug-discovery technologies. He says that doing an industry postdoc is the best way to get introduced to the culture of research and development in the private sector.

A common thread of advice from industry postdocs is that they could not have learned if they were suited for the pharmaceutical industry unless they did a stint in it. Charles Adkins, a senior research biologist at Terlings Park, spent several summers there as an intern during his undergraduate and PhD studies. He says it gave him a valuable insight into the different cultures. "In academia, you have your project and it's your idea, your baby. That's very exciting, but it's also quite solitary. Coming to industry I've found that the main difference is that you have to work as part of a team." But for most who test the waters of private-sector research, that is the big draw. ■

Karen Kreeger is a freelance science writer in Philadelphia.